Noname manuscript No.
(will be inserted by the editor)

Idempotency, output-drivenness and the faithfulness triangle inequality: some consequences of McCarthy's (2003) categoricity generalization

Giorgio Magri

Received: date / Accepted: date

Abstract *Idempotency* requires any phonotactically licit forms to be faithfully realized. *Output-drivenness* requires any discrepancies between underlying and output forms to be driven exclusively by phonotactics. Tesar (2013) and Magri (to appear) provide tight guarantees for OT output-drivenness and idempotency through conditions on the faithfulness constraints. This paper derives analogous faithfulness conditions for HG idempotency and output-drivenness and develops an intuitive interpretation of the various OT and HG faithfulness conditions thus obtained. The intuition is that faithfulness constraints measure the phonological distance between underlying and output forms. They should thus comply with a crucial axiom of the definition of distance, namely that any side of a triangle is shorter than the sum of the other two sides. This intuition leads to a *faithfulness triangle inequality* which is shown to be equivalent to the faithfulness conditions for idempotency and output-drivenness. These equivalences hold under various assumptions, crucially including McCarthy's (2003b) generalization that (faithfulness) constraints are all *categorical*.

The research reported in this paper has been supported by a Marie Curie Intra European Fellowship (Grant agreement number: PIEF-GA-2011-301938) and by an MIT-France seed grant. Parts of this paper have been presented at OCP 12; at WCCFL 33; at the EPG (Experimental Phonology Group) meeting at Utrecht University in June 2015; at the LSA Workshop on Computational Phonology and Morphology at the University of Chicago in July 2015; at the Rutgers Optimality Research Group in September 2015; at MIT in April 2016. Feedback from those audiences is gratefully acknowledged. Finally, I would like to thank Bruce Tesar for very useful comments on earlier versions of this paper.

G. Magri
SFL UMR 7023 (CNRS, University of Paris 8)
61 Rue Pouchet, 75017 Paris, France
E-mail: magrigrg@gmail.com

1 Introduction

This paper looks at two formal grammatical properties that are central to phonological theory and phonological learnability: idempotency and output-drivenness. *Idempotency* formalizes the intuition that any phonotactically licit form is faithfully realized. It has been exploited in the context of phonotactic learning, as it allows the learner to safely assume faithful underlying forms (Prince and Tesar 2004; Hayes 2004). A phonological grammar fails at idempotency provided it displays a *chain shift* (Łubowicz 2011). Failure of idempotency thus corresponds to *counter-feeding opacity.*

Output-drivenness formalizes the intuition that any discrepancy between underlying and surface (or output) forms is driven by the phonotactics. Tesar (2013) shows how to exploit it in the context of the problem of learning underlying forms. Output-drivenness entails idempotency: if repairs are only motivated by the phonotactics (output-drivenness), phonotactically licit forms should be faithfully realized (idempotency). An idempotent grammar fails at output-drivenness provided it displays a *saltation* (White 2013). Failure of output-drivenness thus corresponds to *counter-bleeding opacity.*

Within constraint-based phonology, formal grammatical properties follow from properties of the constraint set.[1] Which assumptions on the constraint set suffice to guarantee that the grammars in the corresponding typology are all idempotent or output-driven? Tesar (2013) tackles this question for output-drivenness within *Optimality Theory* (OT; Prince and Smolensky 2004). Magri (to appear) extends Tesar's analysis to OT idempotency. This paper systematizes those proposals and extends them to *Harmonic Grammar* (HG; Legendre et al 1990b,a; Smolensky and Legendre 2006). Furthermore, it studies the relations among the resulting constraint conditions and relates them to a condition known as the *triangle inequality.* This introductory section offers an informal preview of the main results and their implications.

1.1 Summary of the main contributions

1.1.1 First contribution: more faithfulness constraint conditions

Section 3 reviews Magri's (to appear) theory of OT idempotency. It introduces the *faithfulness idempotency condition* (FIC^{OT}), a condition on the faithfulness constraints which suffices to ensure idempotency of all the grammars in an OT typology. The FIC^{OT} is satisfied by a variety of faithfulness constraints which arise within McCarthy and Prince's (1995) Correspondence Theory and its more recent developments. Section 4 extends the theory of idempotency from OT to HG. Again, HG idempotency is shown to follow from a sufficient condition on the faithfulness constraints, referred to as the FIC^{HG}. In the

[1] Properties of the candidate set also play a crucial role in shaping a typology. This introductory section offers only an informal preview of the main results and thus omits various candidate conditions which will be discussed in the rest of the paper.

general case, the FIC$^{\mathrm{HG}}$ is stronger than the FIC$^{\mathrm{OT}}$. This is expected, given that HG typologies are usually larger than OT typologies, so that a stronger condition is needed to discipline all HG grammars to comply with idempotency.

The rest of the paper turns to output-drivenness. It captures the intuition that, if /red/ is mapped to [rat], /rad/ should be mapped to [rat] as well, because /rad/ is more similar to [rat] than /red/ is. Output-drivenness is thus predicated on a *similarity oder* among input/output mappings. Tesar (2013) offers a *concrete* definition of similarity in terms of strings and correspondence relations. Section 5 reconstructs Tesar's theory using instead an *axiomatic* definition of similarity. The proposed axiom is stated in terms of the faithfulness constraints in an arbitrary faithfulness constraint set $\mathcal{F}$. It captures in particular the intuition that the less similar candidate violates the faithfulness constraints in $\mathcal{F}$ more than the more similar candidate. Tesar's (2013) sufficient faithfulness condition for OT output-drivenness is then reviewed, referred to as the FODC$^{\mathrm{OT}}$. In the general case, the FODC$^{\mathrm{OT}}$ is stronger than the sufficient condition for OT idempotency provided by the FIC$^{\mathrm{OT}}$. This is expected, given that output-drivenness entails idempotency. Finally, Section 6 extends the theory of output-drivenness from OT to HG. Again, HG output-drivenness is shown to follow from a condition on the faithfulness constraints, referred to as the FODC$^{\mathrm{HG}}$. In the general case, the FODC$^{\mathrm{HG}}$ is stronger than both the FODC$^{\mathrm{OT}}$ and the FIC$^{\mathrm{HG}}$.

1.1.2 Second contribution: relating the faithfulness constraint conditions

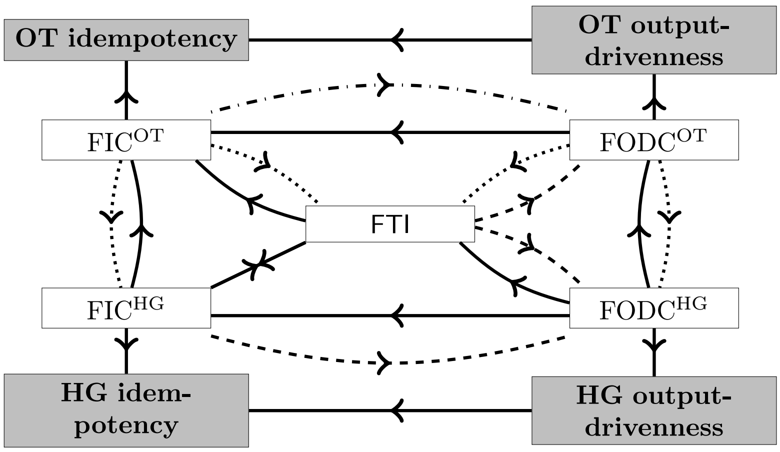
As idempotency and output-drivenness in OT and HG follow from conditions on the faithfulness constraints, the paper digs deeper into the formal underpinning of faithfulness. Faithfulness constraints measure the phonological distance between underlying and surface forms. It thus makes sense to ask whether they satisfy axioms of the abstract definition of *distance*. One such axiom requires the distance between any two points **a** and **c** to be shorter than the distance between **a** and **b** plus the distance between **b** and **c**, for any layover **b**. This axiom is called the *triangle inequality*, as it captures the fact that a side of a triangle is shorter than the sum of the other two sides. These considerations lead to the *faithfulness triangle inequality* (FTI) formulated in Section 2.

Suppose that the faithfulness constraints are all *categorical*, as conjectured in McCarthy (2003b). Section 2 formalizes the notion of categoricity within McCarthy and Prince's (1995) Correspondence Theory of faithfulness. This formalization leads to the main technical result of this paper: Section 3 derives an equivalence result between the sufficient condition for OT idempotency provided by the FIC$^{\mathrm{OT}}$ and the FTI. This equivalence holds for categorical faithfulness constraints, thus distilling a non-trivial implication of McCarthy's categoricity conjecture. The connection between idempotency and the FTI is strengthened in HG: Section 4 shows that the sufficient condition for HG idempotency provided by the FIC$^{\mathrm{HG}}$ is equivalent to the FTI for any faithfulness constraint, independently of its categoricity.

The FTI turns out to play an important role also in the theory of output-drivenness. In fact, Section 5 shows that the FTI entails the sufficient condition for OT output-drivenness provided by Tesar's FODC$^{\mathrm{OT}}$. This entailment holds for any faithfulness constraint which belongs to the selected faithfulness set $\mathcal{F}$ used to measure similarity, independently of its categoricity. For the special case of categorical constraints, the FTI and the FODC$^{\mathrm{OT}}$ turn out to be actually equivalent, as the reverse entailment holds as well (as the FODC$^{\mathrm{OT}}$ is stronger than the FIC$^{\mathrm{OT}}$ which in turn entails the FTI for categorical constraints, as just mentioned). The connection between output-drivenness and the FTI is strengthened in HG: Section 6 shows that the sufficient condition for HG output-drivenness provided by the FODC$^{\mathrm{HG}}$ is equivalent to the FTI for any faithfulness constraint in $\mathcal{F}$, independently of its categoricity.

The resulting network of entailments among the faithfulness conditions considered is summarized in (1). Solid arrows represent entailments which hold for any faithfulness constraint; dotted arrows represent entailments which hold for categorical faithfulness constraints; dashed arrows represent entailments which hold for faithfulness constraints in the faithfulness set $\mathcal{F}$ used to measure similarity; dot-dashed arrows represent entailments which require both restrictions (categoricity and membership in $\mathcal{F}$).

(1)



1.2 Implications

1.2.1 Implications for the interpretation of faithfulness constraint conditions

The sufficient conditions FIC$^{\mathrm{OT}}$, FIC$^{\mathrm{HG}}$, FODC$^{\mathrm{OT}}$, and FODC$^{\mathrm{HG}}$ for OT and HG idempotency and output-drivenness are abstract technical conditions *prima facie* without an intuitive interpretation. How come that most faithfulness constraints in the phonological literature satisfy them (whereby the well known difficulty of modeling opacity in constraint-based phonology)? The FTI at the center of the scheme (1) solves the mystery by providing an intuitive interpretation of these four abstract conditions. According to this interpretation, these conditions simply require faithfulness constraints to measure the phonological distance between underlying and surface forms in compliance with the triangle inequality which characterizes distances. This interpretation holds under no additional assumptions for the HG conditions FIC^{HG} and FODC^{HG} while it requires categoricity for the OT conditions FIC^{OT} and FODC^{OT}.

1.2.2 Implications for the relationship between idempotency and output-drivenness

The sufficient conditions FODC^{OT} and FODC^{HG} for output-drivenness entail the corresponding sufficient conditions FIC^{OT} and FIC^{HG} for idempotency, matching the fact that output-drivenness entails idempotency. Suppose that the faithfulness set $\mathcal{F}$ used to measure similarity in the definition of output-drivenness actually contains all the faithfulness constraints in the constraint set. The scheme (1) then says that the conditions FIC^{OT} and FIC^{HG} for idempotency are equivalent to the corresponding conditions FODC^{OT} and FODC^{HG} for output-drivenness (in the case of OT, the equivalence requires McCarthy's categoricity conjecture). In other words, output-drivenness is stronger than idempotency (in the sense that there exist idempotent grammars which fail at output-drivenness) only if the relation of similarity underlying output-drivenness is blind to some of the faithfulness constraints, namely it is defined relative to a faithfulness set $\mathcal{F}$ which is smaller than or different from the faithfulness constraint set used to define the typology.

1.2.3 Implications for the relationship between OT and HG

The conditions FIC^{HG} and FODC^{HG} for idempotency and output-drivenness in HG entail the corresponding conditions FIC^{OT} and FODC^{OT} in OT, matching the fact that HG typologies are generally larger than OT typologies. Yet, the scheme (1) says that the OT conditions FIC^{OT} and FODC^{OT} are actually equivalent to the corresponding HG conditions FIC^{HG} and FODC^{HG} for faithfulness constraints which are categorical, thus distilling yet another implication of McCarthy's (2003) categoricity conjecture.

1.2.4 Implications for the analysis of faithfulness constraints

Which faithfulness constraints satisfy the conditions FIC^{OT}, FIC^{HG}, FODC^{OT}, or FODC^{HG} for idempotency and output-drivenness in OT and HG? The answer to this question is non-trivial. For instance, two pages of Tesar's book suffice to establish the FODC^{OT} as a sufficient condition for OT output-drivenness, while the entire chapter 3 is devoted to establishing the FODC^{OT} for just the three constraints Max, Dep, and Ident. The scheme (1) affords a substantial simplification. Since the four conditions FIC^{OT}, FIC^{HG}, FODC^{OT}, FODC^{HG} are equivalent for categorical faithfulness constraints, it suffices to focus on the FIC^{OT}, which is arguably the simplest of the four conditions, as revealed by the fact that it is the weakest one in the general case. Magri (to appear) investigates this condition for a large variety of faithfulness constraints which have been proposed within McCarthy and Prince's (1995) Correspondence Theory and its more recent developments. Since those constraints are all categorical, the scheme (1) says that those results concerning the faithfulness constraints which satisfy the FIC^{OT} translate straightforwardly into results concerning the faithfulness constraints which satisfy the other three conditions FIC^{HG}, FODC^{OT}, FODC^{HG}. A measure of the improvement obtained is provided by the fact that a large set of faithfulness constraints (beyond the three considered by Tesar) will be shown in a snap to satisfy the FODC^{OT}.

2 Formal aspects of the theory of faithfulness: triangle inequality and categoricity

This section reviews McCarthy and Prince's (1995) Correspondence Theory of faithfulness (Subsection 2.1) and then focuses on two formal properties of faithfulness constraints: the triangle inequality (Subsection 2.2) and McCarthy's (2003b) categoricity (Subsection 2.3).

2.1 Framework

This Subsection specifies the framework assumed in the paper through some axioms on the candidate and constraint sets.

2.1.1 Candidate set

I assume the representational framework (2), which is a segmental version of McCarthy and Prince's (1995) Correspondence Theory. Underlying and surface forms are strings of segments and phonological candidates establish a correspondence between their segments. Segments are denoted by $\mathsf{a}, \mathsf{b}, \mathsf{c}, \ldots$ and strings by $\mathbf{a}, \mathbf{b}, \mathbf{c}, \ldots$ The notation $\mathbf{a} = \mathsf{a}_1 \cdots \mathsf{a}_\ell$ says that the string $\mathbf{a}$ is the concatenation of the segments $\mathsf{a}_1, \ldots, \mathsf{a}_\ell$ and thus has *length* ℓ.

(2) The *candidate set* consists of triplets $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ of an *underlying* segment string $\mathbf{a}$ and a *surface* segment string $\mathbf{b}$ together with a *correspondence relation* $\rho_{\mathbf{a},\mathbf{b}}$ between the segments of $\mathbf{a}$ and those of $\mathbf{b}$.[2]

Correspondence relations will be denoted by thin lines. To illustrate, (3) represents the candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ whose underlying string $\mathbf{a}$ is /bnɪk/, whose surface string $\mathbf{b}$ is [blɪk], whose correspondence relation $\rho_{\mathbf{a},\mathbf{b}}$ maps underlying to surface segments respecting their position in the strings.

[2] Correspondence relations might want to distinguish between multiple occurrences of the same segment in a string. Thus, correspondence relations cannot be defined simply as relations between the two *sets* of underlying and surface segments. To keep the presentation straightforward, the paper will follow common practice and ignore these subtleties.

(3)
$$\begin{array}{rl} \mathbf{a} = & \text{b n ɪ k} \\ & \;|\;\;|\;\;|\;\;| \\ \mathbf{b} = & \text{b l ɪ k} \end{array}$$

The representational assumption (2) places no *a priori* restrictions neither on the underlying and surface strings nor on the relations which can put them in correspondence. The representational framework is thus compatible with approaches which hardwire restrictions into the candidate set (Blaho et al 2007). I will indeed explore the implications of various restrictions on the correspondence relations which can figure in the candidate set.

2.1.2 Reflexivity and transitivity axioms

As an initial qualification of the representational framework (2), I assume that the candidate set satisfies the *reflexivity* axiom (4): each surface string is required to be in correspondence with itself through the identity correspondence relation which puts each segment in correspondence with itself. This axiom requires each surface form to also count as an underlying form, thus partially blurring the distinction between the two representational levels (for discussion, see Moreton 2004 and Magri to appear, subsection 3.5).

(4) If the candidate set contains a candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ whose surface string is $\mathbf{b}$, then it also contains the *identity* candidate $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$, where $\mathbb{I}_{\mathbf{b},\mathbf{b}}$ is the identity correspondence relation among the segments of $\mathbf{b}$.

Consider two candidates $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ and $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ which share a string $\mathbf{b}$ as the underlying and surface form respectively, as in (5). The *composition* $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$ of the correspondence relations $\rho_{\mathbf{a},\mathbf{b}}$ and $\rho_{\mathbf{b},\mathbf{c}}$ is defined as follows:[3] two segments a and c of the strings $\mathbf{a}$ and $\mathbf{c}$ are in correspondence through $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$ provided there exists a “mediating” segment b of the string $\mathbf{b}$ such that a corresponds to b through $\rho_{\mathbf{a},\mathbf{b}}$ and b corresponds to c through $\rho_{\mathbf{b},\mathbf{c}}$.

(5)
$$\begin{tikzcd} \mathsf{a} \arrow[r, "\rho_{\mathbf{a},\mathbf{b}}"] \arrow[rr, bend right, "\rho_{\mathbf{a},\mathbf{c}} = \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}"'] & \mathsf{b} \arrow[r, "\rho_{\mathbf{b},\mathbf{c}}"] & \mathsf{c} \end{tikzcd}$$

As a further qualification of the representational framework (2), I assume that the candidate set satisfies the *transitivity axiom* (6) which ensures the existence of *composition* candidates.

(6) If the candidate set contains two candidates $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ and $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$, it also contains their *composition candidate* $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$.

Axioms (4) and (6) would hold in particular for an *unrestricted* candidate set which contains all possible correspondence relations.

[3] The operation of composition between two relations is usually denoted by “$\circ$”. In the rest of the paper, I write more succinctly $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$ instead of $\rho_{\mathbf{a},\mathbf{b}} \circ \rho_{\mathbf{b},\mathbf{c}}$.

2.1.3 Constraint set

A *constraint* C takes a candidate $(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}})$ and returns a *number of violations* $C(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}})$ which is large when the candidate scores poorly from the perspective relevant to that constraint. A *markedness* constraint M is defined by the condition (7) that it is blind to underlying forms and thus cannot distinguish between two candidates $(\mathsf{a}, \mathsf{c}, \rho_{\mathsf{a},\mathsf{c}})$ and $(\mathsf{b}, \mathsf{c}, \rho_{\mathsf{b},\mathsf{c}})$ sharing the surface form c.

(7) $M(\mathsf{a}, \mathsf{c}, \rho_{\mathsf{a},\mathsf{c}}) = M(\mathsf{b}, \mathsf{c}, \rho_{\mathsf{b},\mathsf{c}})$

A *faithfulness* constraint F is defined by the condition (8) that it assigns no violations to any identity candidate $(\mathsf{c}, \mathsf{c}, \mathbb{I}_{\mathsf{c},\mathsf{c}})$ in the candidate set.

(8) $F(\mathsf{c}, \mathsf{c}, \mathbb{I}_{\mathsf{c},\mathsf{c}}) = 0$

Suppose that a constraint C is both a faithfulness and a markedness constraint. Consider an arbitrary candidate $(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}})$ in the candidate set. The reflexivity axiom ensures that the candidate set also contains the identity candidate $(\mathsf{b}, \mathsf{b}, \mathbb{I}_{\mathsf{b},\mathsf{b}})$. These two candidates $(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}})$ and $(\mathsf{b}, \mathsf{b}, \mathbb{I}_{\mathsf{b},\mathsf{b}})$ share the surface form b. Since C is a markedness constraint, C must therefore assign them the same number of violations, as in (9a). Since C is also a faithfulness constraint, it does not penalize the identity candidate $(\mathsf{b}, \mathsf{b}, \mathbb{I}_{\mathsf{b},\mathsf{b}})$, as in (9b).

(9) $C(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}}) \overset{(a)}{=} C(\mathsf{b}, \mathsf{b}, \mathbb{I}_{\mathsf{b},\mathsf{b}}) \overset{(b)}{=} 0$

In conclusion, C does not penalize any candidate and it is therefore trivial. This conclusion crucially rests on the reflexivity axiom (4), which ensures that the candidate set has enough identity candidates. Without that axiom, the assumption that C is a faithfulness constraint would indeed have no bite, as the faithfulness definitional condition (8) is stated in terms of identity candidates.

Although no (non-trivial) constraint can be *both* a faithfulness and a markedness constraint, it can be *neither* (*comparative markedness constraints* are one such example; McCarthy 2002, 2003a). To rule out the latter case, I assume throughout the paper Moreton's (2004) *conservativity* assumption (10) that the constraint set only consists of faithfulness and markedness constraints.

(10) constraint set = faithfulness constraints ∪ markedness constraints

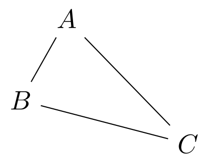
2.2 The faithfulness triangle inequality

This Subsection formalizes the intuition that faithfulness constraints measure the phonological distance between underlying and surface forms in compliance with a core axiom of the notion of distance, namely the *triangle inequality.*

2.2.1 Intuition: the metric triangle inequality

A *distance* (or a *metric*) maps two points A and B to a non-negative value $dist(A, B)$. In order to capture the intuitive notion of distance, this mapping needs to satisfy some core axioms (Rudin 1953, ch. 2). One of these axioms is the *triangle inequality* (11): the distance between two points A and C is never larger than the sum of the distance between A and B plus the distance between B and C, no matter how the intermediate point B is chosen.[4] In other words, no side of a triangle can be longer than the sum of the other two sides.

(11) $dist(A, C) \leq dist(A, B) + dist(B, C)$



Motivated by the intuition that a faithfulness constraint measures the phonological distance between underlying and corresponding surface forms, I extend the metric triangle inequality to faithfulness constraints as condition (12) for any two candidates $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ and $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ which share the string $\mathbf{b}$ as the underlying and the surface form, respectively.[5]

(12) $F\big(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}\big) \leq F\big(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}\big) + F\big(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}\big)$

In order to enforce this condition, we need an explicit assumption on the correspondence relation $\rho_{\mathbf{a},\mathbf{c}}$ which appears on the left-hand side of (12).

2.2.2 Formalization: the FTI$_{comp}$

No faithfulness constraint satisfies condition (12) for *every* correspondence relation $\rho_{\mathbf{a},\mathbf{c}}$, because some choices of $\rho_{\mathbf{a},\mathbf{c}}$ make the left-hand side too large (for instance when $\rho_{\mathbf{a},\mathbf{c}}$ is empty and F is MAX or DEP). Conversely, every faithfulness constraint satisfies condition (12) for *some* correspondence relation $\rho_{\mathbf{a},\mathbf{c}}$, because some choices of $\rho_{\mathbf{a},\mathbf{c}}$ make the left-hand side equal to zero (for instance when $\rho_{\mathbf{a},\mathbf{c}}$ is total and F is MAX or DEP). Some link is needed between the correspondence relations $\rho_{\mathbf{a},\mathbf{b}}$ and $\rho_{\mathbf{b},\mathbf{c}}$ on the right-hand side of (12) and the correspondence relation $\rho_{\mathbf{a},\mathbf{c}}$ on the left-hand side. A natural assumption is that $\rho_{\mathbf{a},\mathbf{c}}$ is the composition $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$ of the correspondence relations $\rho_{\mathbf{a},\mathbf{b}}$ and $\rho_{\mathbf{b},\mathbf{c}}$, whose existence is guaranteed by the transitivity axion (6). In conclusion, a faithfulness constraint F is said to satisfy the *faithfulness triangle inequality* (FTI$_{\text{comp}}$) provided condition (13) holds for any two candidates $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$

[4] Besides the triangle inequality, the abstract definition of distance requires two additional axioms. The first axiom is *symmetry*, which requires the distance between two points to be insensitive to their order. The second axiom is the *identity of the indiscernibles*, which requires two points to coincide if and only if they have zero distance. Together with the triangle inequality, these axioms ensure the non-negativity of a distance. Symmetry fails for core faithfulness constraints such as MAX and DEP. Half of the identity of the indiscernibles is enforced by the definitional condition (8) of faithfulness constraints: if the underlying and surface forms coincide (and the correspondence relation is the identity), their faithfulness violations are equal to zero. But the other half of the identity of the indiscernibles fails, as faithfulness constraints are satisfied by less than perfect string identity. These considerations motivate the focus on the remaining crucial metric axiom, namely the triangle inequality.

[5] The triangle inequality is applied here to each faithfulness constraint individually, rather than to some aggregation (e.g., a weighted sum) of their number of violations.

and $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ and their composition candidate $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$. The subscript "comp" in the acronym FTI_{comp} makes it explicit that the candidate on the left-hand side of the inequality (13) is the composition candidate.

(13) $$F\big(\mathbf{a}, \mathbf{c}, \underbrace{\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}}_{\rho_{\mathbf{a},\mathbf{c}}}\big) \leq F\big(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}\big) + F\big(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}\big)$$

Subsection 3.3.5 will show that a wide range of faithfulness constraints within Correspondence Theory satisfy the FTI_{comp}.

2.3 Faithfulness categoricity, additivity and monotonicity

McCarthy (2003b) conjectures that phonological constraints are all categorical. Informally, this means that they assign at most one violation per *locus* of violation. McCarthy provides an explicit formalization of categoritcity for markedness constraints (see his scheme (1) on p. 77). His treatment of faithfulness constraints is not as explicit: he discusses individual faithfulness constraints but does not provide a general scheme. This Subsection fills the gap. The idea of the formalization is that a candidate can be sliced into "sub-candidates" by slicing either its correspondence relation, or its underlying string, or its surface string. A faithfulness constraint is additive provided the number of violations it assigns to a candidate is the sum of the number of violations it assigns to the sub-candidates. An additive faithfulness constraint is categorical provided it only takes values 0 or 1 when applied to the sub-candidates. Finally, additivity entails monotonicity, which says that the number of violations of bigger candidates (namely, those with more sub-candidates) ought to be larger.

2.3.1 Additivity, categoricity, and monotonicity w.r.t. correspondence relations

Intuitively, the identity faithfulness constraint $\textsc{Ident}_\varphi$ corresponding to a (total) feature φ counts over pairs of corresponding segments (McCarthy and Prince 1995). This intuition can be formalized through the following two-step definition. First, the constraint is defined for a candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ whose correspondence relation $\rho_{\mathbf{a},\mathbf{b}}$ consists of a single segment pair (a, b), as in (14a). Second, the definition is extended to an arbitrary candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ by summing over pairs (a, b) in the correspondence relation $\rho_{\mathbf{a},\mathbf{b}}$, as in (14b).

(14) a. $$\textsc{Ident}_\varphi\big(\mathbf{a}, \mathbf{b}, \{(\mathsf{a}, \mathsf{b})\}\big) = \begin{cases} 1 & \text{if } \mathsf{a} \text{ and } \mathsf{b} \text{ differ relative to feature } \varphi \\ 0 & \text{otherwise} \end{cases}$$

b. $$\textsc{Ident}_\varphi\big(\mathbf{a},\ \mathbf{b},\ \rho_{\mathbf{a},\mathbf{b}}\big) = \sum_{(\mathsf{a},\mathsf{b}) \in \rho_{\mathbf{a},\mathbf{b}}} \textsc{Ident}_\varphi\big(\mathbf{a},\ \mathbf{b},\ \{(\mathsf{a}, \mathsf{b})\}\big)$$

Intuitively, the faithfulness constraints $\textsc{Linearity}$ (McCarthy and Prince 1995) and I/O-$\textsc{Adjacency}$ (Carpenter 2002) count over two pairs of corresponding segments. This intuition can be formalized through the following two-step definition. First, the constraints are defined for a candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ whose correspondence relation $\rho_{\mathbf{a},\mathbf{b}}$ consists of only two pairs of segments $(\mathsf{a}_1, \mathsf{b}_1)$ and $(\mathsf{a}_2, \mathsf{b}_2)$, as in (15). I also consider the variant I/O-ADJACENCY$^{\text{grad}}$ in (16), which is sensitive to the number of intervening segments.

(15) $$\textsc{Linearity}\big(\mathbf{a},\ \mathbf{b},\ \{(\mathsf{a}_1, \mathsf{b}_1), (\mathsf{a}_2, \mathsf{b}_2)\}\big) = \\ = \begin{cases} 1 & \mathsf{a}_1 \text{ precedes } \mathsf{a}_2 \text{ in } \mathbf{a} \text{ but } \mathsf{b}_1 \text{ follows } \mathsf{b}_2 \text{ in } \mathbf{b} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{I-}\textsc{Adjacency}\big(\mathbf{a},\ \mathbf{b},\ \{(\mathsf{a}_1, \mathsf{b}_1), (\mathsf{a}_2, \mathsf{b}_2)\}\big) = \\ = \begin{cases} 1 & \text{if } \mathsf{a}_1, \mathsf{a}_2 \text{ are adjacent in } \mathbf{a} \text{ but } \mathsf{b}_1, \mathsf{b}_2 \text{ are not adjacent in } \mathbf{b} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{O-}\textsc{Adjacency}\big(\mathbf{a},\ \mathbf{b},\ \{(\mathsf{a}_1, \mathsf{b}_1), (\mathsf{a}_2, \mathsf{b}_2)\}\big) = \\ = \begin{cases} 1 & \text{if } \mathsf{b}_1, \mathsf{b}_2 \text{ are adjacent in } \mathbf{b} \text{ but } \mathsf{a}_1, \mathsf{a}_2 \text{ are not adjacent in } \mathbf{a} \\ 0 & \text{otherwise} \end{cases}$$

(16) $$\text{I-}\textsc{Adjacency}^{\text{grad}}\big(\mathbf{a},\ \mathbf{b},\ \{(\mathsf{a}_1, \mathsf{b}_1), (\mathsf{a}_2, \mathsf{b}_2)\}\big) = \\ = \begin{cases} k & \text{if } \mathsf{a}_1, \mathsf{a}_2 \text{ are adjacent but } \mathsf{b}_1, \mathsf{b}_2 \text{ are separated by } k \text{ segments} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{O-}\textsc{Adjacency}^{\text{grad}}\big(\mathbf{a},\ \mathbf{b},\ \{(\mathsf{a}_1, \mathsf{b}_1), (\mathsf{a}_2, \mathsf{b}_2)\}\big) = \\ = \begin{cases} k & \text{if } \mathsf{b}_1, \mathsf{b}_2 \text{ are adjacent but } \mathsf{a}_1, \mathsf{a}_2 \text{ are separated by } k \text{ segments} \\ 0 & \text{otherwise} \end{cases}$$

Second, the constraints are defined for an arbitrary candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ by summing over any two pairs $(\mathsf{a}_1, \mathsf{b}_1)$ and $(\mathsf{a}_2, \mathsf{b}_2)$ in the relation $\rho_{\mathbf{a},\mathbf{b}}$, as in (17).

(17) $$F\big(\mathbf{a},\ \mathbf{b},\ \rho_{\mathbf{a},\mathbf{b}}\big) = \sum_{(\mathsf{a}_1,\mathsf{b}_1),(\mathsf{a}_2,\mathsf{b}_2) \in \rho_{\mathbf{a},\mathbf{b}}} F\big(\mathbf{a},\ \mathbf{b},\ \{(\mathsf{a}_1, \mathsf{b}_1), (\mathsf{a}_2, \mathsf{b}_2)\}\big)$$

Generalizing from these examples, a faithfulness constraint F is called *additive relative to the correspondence relations* (C-additive) of order ℓ provided the identity (18) holds for any candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$. The sum is over subsets of cardinality ℓ of the correspondence relation $\rho_{\mathbf{a},\mathbf{b}}$ (not necessarily disjunct one from the other). In other words, the number of violations assigned by F to the candidate $(\mathbf{a},\ \mathbf{b},\ \rho_{\mathbf{a},\mathbf{b}})$ is the sum of the number of violations that F assigns to the "sub-candidates" $(\mathbf{a},\ \mathbf{b},\ \{(\mathsf{a}_1, \mathsf{b}_1), \ldots, (\mathsf{a}_\ell, \mathsf{b}_\ell)\})$ obtained by considering all correspondence sub-relations $\{(\mathsf{a}_1, \mathsf{b}_1), \ldots, (\mathsf{a}_\ell, \mathsf{b}_\ell)\}$ of cardinality ℓ of $\rho_{\mathbf{a},\mathbf{b}}$.

(18) $$F\big(\mathbf{a},\ \mathbf{b},\ \rho_{\mathbf{a},\mathbf{b}}\big) = \sum_{(\mathsf{a}_1,\mathsf{b}_1),\ldots,(\mathsf{a}_\ell,\mathsf{b}_\ell) \in \rho_{\mathbf{a},\mathbf{b}}} F\big(\mathbf{a},\ \mathbf{b},\ \{(\mathsf{a}_1, \mathsf{b}_1), \ldots, (\mathsf{a}_\ell, \mathsf{b}_\ell)\}\big)$$

F is called *categorical relative to the correspondence relations* (C-categorical) of order ℓ provided it is additive and furthermore can only take values 0 or 1 when it is applied to a candidate whose correspondence relation has cardinality ℓ, so that the addenda on the right-hand side of (18) are all equal to 0 or 1. To illustrate, the constraint IDENT$_\varphi$ is C-categorical of order 1; the constraints I/O-ADJACENCY and LINEARITY are C-categorical of order 2; the constraints I/O-ADJACENCY$^{\text{grad}}$ are C-additive but not C-categorical.

Finally, a faithfulness constraint F is called *monotone relative to the correspondence relations* (C-monotone) provided it satisfies the implication (19): if two candidates $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ and $(\mathbf{a}, \mathbf{b}, \widehat{\rho}_{\mathbf{a},\mathbf{b}})$ share both the underlying string $\mathbf{a}$ and the surface string $\mathbf{b}$ and only differ because the correspondence relation of the former is a subset of the correspondence relation of the latter, F assigns less violations to the former candidate than to the latter.

(19) **If:** $\rho_{\mathbf{a},\mathbf{b}} \subseteq \widehat{\rho}_{\mathbf{a},\mathbf{b}}$
Then: $F\big(\mathbf{a},\, \mathbf{b},\, \rho_{\mathbf{a},\mathbf{b}}\big) \leq F\big(\mathbf{a},\, \mathbf{b},\, \widehat{\rho}_{\mathbf{a},\mathbf{b}}\big)$

Any C-additive faithfulness constraint is C-monotone, as a larger correspondence relation yields more addenda on the right-hand side of (18).

2.3.2 Additivity, categoricity, and monotonicity w.r.t. underlying strings

Intuitively, the faithfulness constraints MAX and INTEGRITY (McCarthy and Prince 1995), INTEGRITY$^{\text{grad}}$ (Wheeler 2005) and MAX$_{[+\varphi]}$ (Casali 1997, 1998; Walker 1999; Lombardi 2001) count over underlying segments. This intuition can be formalized through the following two-step definition. First, these constraints are defined for a candidate $(\mathsf{a}, \mathbf{b}, \rho_{\mathsf{a},\mathbf{b}})$ whose underlying string consists of a single segment a, as in (20).

(20)
$$\textsc{Max}\big(\mathsf{a}, \mathbf{b}, \rho_{\mathsf{a},\mathbf{b}}\big) = \begin{cases} 1 & \text{if } \mathsf{a} \text{ has no surface correspondent w.r.t } \rho_{\mathsf{a},\mathbf{b}} \\ 0 & \text{otherwise} \end{cases}$$

$$\textsc{Int}\big(\mathsf{a}, \mathbf{b}, \rho_{\mathsf{a},\mathbf{b}}\big) = \begin{cases} 1 & \text{if } \mathsf{a} \text{ has multiple correspondents w.r.t. } \rho_{\mathsf{a},\mathbf{b}} \\ 0 & \text{otherwise} \end{cases}$$

$$\textsc{Int}^{\text{grad}}\big(\mathsf{a}, \mathbf{b}, \rho_{\mathsf{a},\mathbf{b}}\big) = \begin{cases} k & \text{if } \mathsf{a} \text{ has } k \geq 2 \text{ correspondents w.r.t. } \rho_{\mathsf{a},\mathbf{b}} \\ 0 & \text{otherwise} \end{cases}$$

$$\textsc{Max}_{[+\varphi]}\big(\mathsf{a}, \mathbf{b}, \rho_{\mathsf{a},\mathbf{b}}\big) = \begin{cases} 1 & \text{if } \mathsf{a} \text{ has value + for feature } \varphi \text{ but } \mathsf{a} \text{ has no correspondent } \mathsf{b} \text{ w.r.t } \rho_{\mathsf{a},\mathbf{b}} \text{ such that } \mathsf{b} \text{ also has value + for feature } \varphi \\ 0 & \text{otherwise} \end{cases}$$

Second, the constraints are defined for a generic candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ by summing over each segment a of the underlying string $\mathbf{a}$, as in (21).[6] The correspondence relation $\rho_{\mathbf{a},\mathbf{b}} \restriction_{(\mathsf{a},\mathbf{b})}$ on the right-hand side of (21) is the restriction of the correspondence relation $\rho_{\mathbf{a},\mathbf{b}}$ from the string $\mathbf{a}$ to its segment a.[7]

(21)
$$F\big(\mathbf{a},\, \mathbf{b},\, \rho_{\mathbf{a},\mathbf{b}}\big) = \sum_{\mathsf{a} \subseteq \mathbf{a}} F\big(\mathsf{a},\, \mathbf{b},\, \rho_{\mathbf{a},\mathbf{b}} \restriction_{(\mathsf{a},\mathbf{b})}\big)$$

[6] For any two strings $\mathbf{a}$ and $\mathbf{b}$, the notation $\mathbf{a} \subseteq \mathbf{b}$ indicates that $\mathbf{a}$ is a subsequence of $\mathbf{b}$: $\mathbf{a}$ is obtained from $\mathbf{b}$ by replacing some symbols of $\mathbf{b}$ with the empty symbol.

[7] In other words, $\rho_{\mathbf{a},\mathbf{b}} \restriction_{(\mathsf{a},\mathbf{b})}$ is the set of those pairs $(\mathsf{a}', \mathsf{b}')$ in $\rho_{\mathbf{a},\mathbf{b}}$ such that $\mathsf{a}' = \mathsf{a}$. Once the underlying string $\mathbf{a}$ is restricted to a single segment a, the correspondence relation $\rho_{\mathbf{a},\mathbf{b}}$ must necessarily be restricted to $\rho_{\mathbf{a},\mathbf{b}} \restriction_{(\mathsf{a},\mathbf{b})}$. In fact, the triplet $(\mathsf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ (with the singleton underlying segment a and the original correspondence relation $\rho_{\mathbf{a},\mathbf{b}}$) does not count as a candidate according to the assumption (2) that correspondence relations hold between the segments of the two strings in the candidate.

Intuitively, the faithfulness constraint MaxLinearity (Heinz 2005) counts over subsequences of length 2 of the underlying string. This intuition can be formalized through the following two-step definition. First, the constraint is defined for a candidate $(\mathsf{a}_1\mathsf{a}_2, \mathbf{b}, \rho_{\mathsf{a}_1\mathsf{a}_2,\mathbf{b}})$ whose underlying string $\mathsf{a}_1\mathsf{a}_2$ has length 2, as in (22a). Second, the constraint is defined for an arbitrary candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ by summing over all subsequences $\mathsf{a}_1\mathsf{a}_2$ of length 2 of the underlying string $\mathbf{a}$, as in (22b).

(22) a. $$\textsc{MaxLin}\big(\mathsf{a}_1\mathsf{a}_2,\, \mathbf{b},\, \rho_{\mathsf{a}_1\mathsf{a}_2,\mathbf{b}}\big) = \begin{cases} 1 & \text{if } \mathsf{a}_1, \mathsf{a}_2 \text{ have no correspondents} \\ & \mathsf{b}_1, \mathsf{b}_2 \text{ such that } \mathsf{b}_1 \text{ precedes } \mathsf{b}_2 \\ 0 & \text{otherwise} \end{cases}$$

b. $$\textsc{MaxLin}\big(\mathbf{a},\, \mathbf{b},\, \rho_{\mathbf{a},\mathbf{b}}\big) = \sum_{\mathsf{a}_1\mathsf{a}_2 \subseteq \mathbf{a}} \textsc{MaxLin}\big(\mathsf{a}_1\mathsf{a}_2,\, \mathbf{b},\, \rho_{\mathbf{a},\mathbf{b}} \restriction_{(\mathsf{a}_1\mathsf{a}_2,\mathbf{b})}\big)$$

Generalizing from these examples, a faithfulness constraint F is called *additive relative to the underlying strings* (I-additive) of order ℓ provided the identity (23) holds for any candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$. The sum is over all (possibly overlapping) subsequences of $\mathbf{a}$ of length ℓ.[8]

(23) $$F\big(\mathbf{a},\, \mathbf{b},\, \rho_{\mathbf{a},\mathbf{b}}\big) = \sum_{\mathsf{a}_1 \cdots \mathsf{a}_\ell \subseteq \mathbf{a}} F\big(\mathsf{a}_1 \cdots \mathsf{a}_\ell,\, \mathbf{b},\, \rho_{\mathbf{a},\mathbf{b}} \restriction_{(\mathsf{a}_1 \cdots \mathsf{a}_\ell,\mathbf{b})}\big)$$

F is called *categorical relative to the underlying strings* (I-categorical) of order ℓ provided it is additive and furthermore it can only take values 0 or 1 when it is applied to a candidate whose underlying string has length ℓ, so that the addenda on the right-hand side of (23) are all equal to 0 or 1. To illustrate, the constraints Max, Integrity, and $\textsc{Max}_{[+\varphi]}$ are I-categorical of order 1; the constraint MaxLinearity is I-categorical of order 2; and the constraint $\textsc{Integrity}^{\text{grad}}$ is I-additive but not I-categorical.

Finally, a faithfulness constraint F is called *monotone relative to the underlying strings* (I-monotone) provided it satisfies the implication (24): if two candidates $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ and $(\widehat{\mathbf{a}}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ share both the surface string $\mathbf{b}$ and the correspondence relation $\rho_{\mathbf{a},\mathbf{b}}$ and only differ because the underlying string $\mathbf{a}$ of the former is a subsequence of the underlying string $\widehat{\mathbf{a}}$ of the latter,[9] then F assigns less violations to the former candidate than to the latter.

[8] A *subsequence* needs not consist of contiguous elements, contrary to a *substring*: itk is both a subsequence and a substring of pitkol, while ptkl is a subsequence but not a substring. It might be possible to define I-additivity in terms of a sum over *sub-strings* of contiguous segments, rather than over *sub-sequences* of possibly non-contiguous strings. For Linearity-type constraints, this modification would capture Heinz's (2005) proposal that only *immediate* precedence matters in the definition of the faithfulness constraints. Switching from sub-sequences to sub-strings would also have implications for Adjacency-type constraints. They have been defined in (15) and (17) as counting over corresponding pairs, whereby they qualify as C-categorical. An alternative definition of, say, I-Adjacency would be the following: it assigns to a candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ a number of violations equal to the number of underling adjacent pairs of segments which have no adjacent surface correspondents. If I-additivity is redefined in terms of sub-strings, then I-Adjacency qualifies as I-additive (it would not count as I-additive according to definition in terms of sub-sequences).

[9] This implies that the shared correspondence relation must hold between the shared surface string $\mathbf{b}$ and the "smaller" underlying string $\mathbf{a}$.

(24) **If:** $\mathsf{a} \subseteq \widehat{\mathsf{a}}$
Then: $F\big(\mathsf{a},\, \mathsf{b},\, \rho_{\mathsf{a},\mathsf{b}}\big) \leq F\big(\widehat{\mathsf{a}},\, \mathsf{b},\, \rho_{\mathsf{a},\mathsf{b}}\big)$

Any I-additive faithfulness constraint is I-monotone, as a longer underlying string yields more addenda on the right-hand side of (23).

2.3.3 A remark on the proper definition of I-additivity

Let me entertain an alternative definition of I-additivity through the alternative condition (25) instead of the actual condition (23) considered above. In both the actual condition (23) and the variant (25), the correspondence relation $\rho_{\mathsf{a},\mathsf{b}}$ which appears on the left-hand side is replaced on the right-hand side with its restriction $\rho_{\mathsf{a},\mathsf{b}}\restriction_{(\mathsf{a}_1\cdots\mathsf{a}_\ell,\mathsf{b})}$ to the underlying subsequence $\mathsf{a}_1\cdots\mathsf{a}_\ell$. Yet, the actual condition (23) and the variant (25) differ because only the former also replaces the underlying string a with the underlying subsequence $\mathsf{a}_1\cdots\mathsf{a}_\ell$. The variant (25) thus makes I-additivity completely analogous to condition (18) used to define C-additivity: in both cases, the right-hand side sums over candidates which share the underlying and surface strings of the original candidate, but have a restricted correspondence relation.

$$(25) \quad F(\mathsf{a},\, \mathsf{b},\, \rho_{\mathsf{a},\mathsf{b}}) = \sum_{\mathsf{a}_1\cdots\mathsf{a}_\ell\subseteq\mathsf{a}} F\big(\mathsf{a},\, \mathsf{b},\, \rho_{\mathsf{a},\mathsf{b}}\restriction_{(\mathsf{a}_1\cdots\mathsf{a}_\ell,\mathsf{b})}\big)$$

Yet, this alternative condition (25) makes no sense. To see that, suppose that the original correspondence relation is empty: $\rho_{\mathsf{a},\mathsf{b}} = \emptyset$. Its restriction to any underlying subsequence is thus empty as well: $\rho_{\mathsf{a},\mathsf{b}}\restriction_{(\mathsf{a}_1\cdots\mathsf{a}_\ell,\mathsf{b})} = \emptyset$. Since the original and the restricted correspondence relations thus coincide, the original candidate $(\mathsf{a},\, \mathsf{b},\, \rho_{\mathsf{a},\mathsf{b}})$ which appears on the left-hand side of (25) and the candidates $(\mathsf{a},\, \mathsf{b},\, \rho_{\mathsf{a},\mathsf{b}}\restriction_{(\mathsf{a}_1\cdots\mathsf{a}_\ell,\mathsf{b})})$ which appear on the right-hand side coincide and must therefore be assigned the same number of violations by the faithfulness constraint F. Condition (25) is thus contradictory, unless F never assigns any violations in the case of empty correspondence relations (which is obviously false for instance in the case of Max).

2.3.4 Additivity, categoricity, and monotonicity w.r.t. surface strings

Intuitively, the faithfulness constraints Dep, Uniformity, Uniformity$^{\text{grad}}$, and Dep$_{[+\varphi]}$ count over surface segments. As in Subsection 2.3.2, this intuition can be formalized through the following two-step definition. First, the constraints are defined for a candidate $(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}})$ whose surface string consists of a single segment b, as in (26).

(26) $$\textsc{Dep}\big(\mathbf{a}, \mathsf{b}, \rho_{\mathbf{a},\mathsf{b}}\big) = \begin{cases} 1 & \text{if } \mathsf{b} \text{ has no underlying correspondent w.r.t } \rho_{\mathbf{a},\mathsf{b}} \\ 0 & \text{otherwise} \end{cases}$$

$$\textsc{Unif}\big(\mathbf{a}, \mathsf{b}, \rho_{\mathbf{a},\mathsf{b}}\big) = \begin{cases} 1 & \text{if } \mathsf{b} \text{ has multiple correspondents w.r.t. } \rho_{\mathbf{a},\mathsf{b}} \\ 0 & \text{otherwise} \end{cases}$$

$$\textsc{Unif}^{\mathrm{grad}}\big(\mathbf{a}, \mathsf{b}, \rho_{\mathbf{a},\mathsf{b}}\big) = \begin{cases} k & \text{if } \mathsf{b} \text{ has } k \geq 2 \text{ correspondents w.r.t. } \rho_{\mathbf{a},\mathsf{b}} \\ 0 & \text{otherwise} \end{cases}$$

$$\textsc{Dep}_{[+\varphi]}\big(\mathbf{a}, \mathsf{b}, \rho_{\mathbf{a},\mathsf{b}}\big) = \begin{cases} 1 & \text{if } \mathsf{b} \text{ has value } + \text{ for feature } \varphi \text{ but } \mathsf{b} \text{ has no correspondent } \mathsf{a} \text{ w.r.t } \rho_{\mathbf{a},\mathsf{b}} \text{ such that } \mathsf{a} \text{ also has value } + \text{ for feature } \varphi \\ 0 & \text{otherwise} \end{cases}$$

Second, the constraints are defined for a generic candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ by summing over the segments b of the surface string $\mathbf{b}$, as in (27). The correspondence relation $\rho_{\mathbf{a},\mathbf{b}}\restriction_{(\mathbf{a},\mathsf{b})}$ which appears on the right-hand side of (27) is the restriction of the original correspondence relation $\rho_{\mathbf{a},\mathbf{b}}$ to the surface segment b.

(27) $$F\big(\mathbf{a},\, \mathbf{b},\, \rho_{\mathbf{a},\mathbf{b}}\big) = \sum_{\mathsf{b} \subseteq \mathbf{b}} F\big(\mathbf{a},\, \mathsf{b},\, \rho_{\mathbf{a},\mathbf{b}}\restriction_{(\mathbf{a},\mathsf{b})}\big)$$

Intuitively, the faithfulness constraint DepLinearity (Heinz 2005) counts over subsequences of length 2 of the surface string. This intuition can be formalized through the following two-step definition. First, the constraint is defined for a candidate $(\mathbf{a}, \mathsf{b}_1\mathsf{b}_2, \rho_{\mathbf{a},\mathsf{b}_1\mathsf{b}_2})$ whose surface string $\mathsf{b}_1\mathsf{b}_2$ has length 2, as in (28a). Second, the constraint is defined for an arbitrary candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ by summing over all surface subsequences $\mathsf{b}_1\mathsf{b}_2$ of length 2, as in (28b).

(28) a. $$\textsc{DepLin}\big(\mathbf{a},\, \mathsf{b}_1\mathsf{b}_2,\, \rho_{\mathbf{a},\mathsf{b}_1\mathsf{b}_2}\big) = \begin{cases} 1 & \text{if } \mathsf{b}_1, \mathsf{b}_2 \text{ have no correspondents } \mathsf{a}_1, \mathsf{a}_2 \text{ such that } \mathsf{a}_1 \text{ precedes } \mathsf{a}_2 \\ 0 & \text{otherwise} \end{cases}$$

b. $$\textsc{DepLin}\big(\mathbf{a},\, \mathbf{b},\, \rho_{\mathbf{a},\mathbf{b}}\big) = \sum_{\mathsf{b}_1\mathsf{b}_2 \subseteq \mathbf{b}} \textsc{DepLin}\big(\mathbf{a},\, \mathsf{b}_1\mathsf{b}_2,\, \rho_{\mathbf{a},\mathbf{b}}\restriction_{(\mathbf{a},\mathsf{b}_1\mathsf{b}_2)}\big)$$

Generalizing from these examples, a faithfulness constraint F is called *additive relative to the surface strings* (O-additive) of order ℓ provided the identity (29) holds for any candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$. The sum is over all (possibly overlapping) subsequences of $\mathbf{b}$ of length ℓ.

(29) $$F\big(\mathbf{a},\, \mathbf{b},\, \rho_{\mathbf{a},\mathbf{b}}\big) = \sum_{\mathsf{b}_1 \cdots \mathsf{b}_\ell \subseteq \mathbf{b}} F\big(\mathbf{a},\, \mathsf{b}_1 \cdots \mathsf{b}_\ell,\, \rho_{\mathbf{a},\mathbf{b}}\restriction_{(\mathbf{a},\mathsf{b}_1\cdots\mathsf{b}_\ell)}\big)$$

F is called *categorical relative to the surface strings* (O-categorical) of order ℓ provided it is additive and furthermore it can only take values 0 or 1 when it is applied to a candidate whose surface string has length ℓ, so that the addenda on the right-hand side of (29) are all equal to 0 or 1. To illustrate, the constraints Dep, Uniformity, and $\textsc{Dep}_{[+\varphi]}$ are O-categorical of order 1; the constraint DepLinearity is O-categorical of order 2; and the constraint $\textsc{Uniformity}^{\mathrm{grad}}$ is O-additive but not O-categorical.

Finally, a faithfulness constraint F is called *monotone relative to the surface strings* (O-monotone) provided it satisfies the implication (30): if two candidates $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ and $(\mathbf{a}, \widehat{\mathbf{b}}, \rho_{\mathbf{a},\mathbf{b}})$ share both the underlying string $\mathbf{a}$ and the correspondence relation $\rho_{\mathbf{a},\mathbf{b}}$ and only differ because the surface string $\mathbf{b}$ of the former is a subsequence of the surface string $\widehat{\mathbf{b}}$ of the latter, then F assigns less violations to the former candidate than to the latter.

(30) **If:** $\mathbf{b} \subseteq \widehat{\mathbf{b}}$
Then: $F\big(\mathbf{a},\, \mathbf{b},\, \rho_{\mathbf{a},\mathbf{b}}\big) \leq F\big(\mathbf{a},\, \widehat{\mathbf{b}},\, \rho_{\mathbf{a},\mathbf{b}}\big)$

Any O-additive faithfulness constraint is O-monotone, as a longer surface string yields more addenda on the right-hand side of (29).

2.3.5 McCarthy's (strengthened) categoricity conjecture

McCarthy (2003b) conjectures that every constraint which is relevant for phonological theory is categorical. Given the preceding discussion, this conjecture takes the following form for the faithfulness constraints.

McCarthy's categoricity conjecture. *Any faithfulness constraint which is relevant for phonological theory is either C-categorical, or I-categorical or O-categorical.* □

The notion of categoricity builds on the notion of additivity. As remarked in Subsection 2.3.3, there is a slight formal asymmetry between C-additivity on the one hand and I-/O-additivity on the other hand: all three notions of additivity require a restriction of the correspondence relations; yet, only I-/O-additivity (but not C-additivity) requires additional restrictions on the underlying and surface strings. I will now formulate a slightly stronger version of McCarthy's categoricity conjecture, which takes into account this asymmetry.

The constraints Max, $\text{Max}_{[+\varphi]}$, Integrity, and MaxLinearity are I-categorical and therefore automatically I-monotone, as shown in Subsection 2.3.2. They are also O-monotone: the number of violations does not shrink (actually, it does not change at all) when the surface string is extended with additional segments (while keeping the correspondence relation unchanged). This is illustrated in (31) for Max: the additional surface segment [i] in $\widehat{\mathbf{b}}$ does not affect the number of deleted underlying segments (the correspondence relation is the same in the two candidates, as required by O-monotonicity).

(31) a. $\text{Max}\big(\mathbf{a},\, \mathbf{b},\, \rho_{\mathbf{a},\mathbf{b}}\big) = 1$

$\mathbf{a} =$ u k a r

$\mathbf{b} =$ k r a

b. $\text{Max}\big(\mathbf{a},\, \widehat{\mathbf{b}},\, \rho_{\mathbf{a},\mathbf{b}}\big) = 1$

$\mathbf{a} =$ u k a r

$\widehat{\mathbf{b}} =$ k i r a

Analogously, the constraints Dep, $\text{Dep}_{[+\varphi]}$, Uniformity, and DepLinearity are O-categorical and therefore automatically O-monotone, as shown in Subsection 2.3.4. They are also I-monotone: the number of violations does not shrink (actually, it does not change at all) when the underlying string is extended with additional segments. These considerations suggest that McCarthy's categoricity conjecture can be strengthened as follows, complementing categoricity with monotonicity. This revised formulation of the categoricity conjecture captures the formal asymmetry between C-categoricity and I-/O-categoricity. This revision is crucial for proposition 3 to hold.

Strengthened categoricity conjecture. *Any faithfulness constraint which is relevant for phonological theory is either C-categorical; or I-categorical* and *O-monotone; or O-categorical* and *I-monotone.* □

The conjecture makes no assumptions on the monotonicity of C-categorical constraints. Indeed, the C-categorical constraint I-Adjacency (O-Adjacency) is O-monotone but not I-monotone (I-monotone but not O-monotone), as additional underlying segments can disrupt adjacency and thus reduce violations.[10]

2.4 Summary

This section has introduced formal conditions on the faithfulness constraints in Correspondence theory. The triangle inequality (formalized as the $\mathrm{FTI}_{\mathrm{comp}}$) captures the intuition that faithfulness constraints measure the phonological distance between underlying and surface forms in compliance with a crucial axiom of the notion of distance. Additivity says that the number of violations assigned to a candidate is the sum of the number of violations assigned to the sub-candidates obtained by slicing the original candidate along any of its three ingredients. Categoricity says that these sub-candidates can be assigned only 0 or 1 violation. Finally, additivity is closely related to monotonicity, which says that the number of violations of bigger candidates (namely, those with more sub-candidates) ought to be larger. The rest of the paper explores the implications of these formal notions for the theory of idempotency and output-drivenness in constraint-based phonology.

3 Idempotency in Optimality Theory

This Section recalls the structural typological condition of idempotency and shows that OT idempotency is related to the faithfulness triangle inequality under McCarthy's (strengthened) categoricity conjecture.

3.1 Idempotency

Within the representational framework defined in Section 2, a phonological grammar G maps an underlying form $\mathbf{a}$ to a candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ whose un-

[10] The alternative definition of I-Adjacency in footnote 8 makes it I-additive. Crucially, it is also O-monotone: adding surface segments can only disrupt surface adjacency and thus increase the number of violations. Analogous considerations hold for O-Adjacency.

derlying string is indeed a.[11] A grammar G is called *idempotent* provided it maps any phonotactically licit form to itself, as formalized in the implication (32). The antecedent of the implication says that the surface form b is *phonotactically licit* relative to the grammar G, because G maps some underlying form a to a candidate $(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}})$ whose surface string is b. The consequent says that b is mapped faithfully to itself. The reflexivity axiom (4) ensures the existence of the identity candidate $(\mathsf{b}, \mathsf{b}, \mathbb{I}_{\mathsf{b},\mathsf{b}})$ in the consequent of (32).

Definition 1 (Idempotency) *A grammar G is* idempotent *provided it satisfies the following implication*

(32) **If:** $G(\boldsymbol{a}) = \big(\boldsymbol{a}, \boldsymbol{b}, \rho_{\boldsymbol{a},\boldsymbol{b}}\big)$
Then: $G(\boldsymbol{b}) = \big(\boldsymbol{b}, \boldsymbol{b}, \mathbb{I}_{\boldsymbol{b},\boldsymbol{b}}\big)$

for any candidate $(\boldsymbol{a}, \boldsymbol{b}, \rho_{\boldsymbol{a},\boldsymbol{b}})$ in the candidate set. □

To illustrate, suppose that a grammar enforces final devoicing and thus maps the underlying form a = /rad/ to b = [rat]. The surface form b = [rat] is thus phonotactically licit. Idempotency thus requires the underlying form b = /rat/ to be faithfully mapped to b = [rat]. Failure of idempotency corresponds to chain-shifts and thus captures counter-feeding opacity.

3.2 Candidate and constraint conditions for OT idempotency

This Section briefly reviews Magri's (to appear) theory of idempotency in OT.

3.2.1 Optimality Theory

A constraint C *prefers* a candidate $(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}})$ to another candidate $(\mathsf{c}, \mathsf{d}, \rho_{\mathsf{c},\mathsf{d}})$ provided it assigns less violations to the former than to the latter, namely $C(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}}) < C(\mathsf{c}, \mathsf{d}, \rho_{\mathsf{c},\mathsf{d}})$. A *constraint ranking* is an arbitrary linear order $\gg$ over a set of constraints. A constraint ranking $\gg$ *prefers* a candidate $(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}})$ to another candidate $(\mathsf{c}, \mathsf{d}, \rho_{\mathsf{c},\mathsf{d}})$ provided the $\gg$-highest constraint which prefers one of two candidates $(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}})$ and $(\mathsf{c}, \mathsf{d}, \rho_{\mathsf{c},\mathsf{d}})$, prefers in particular $(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}})$. The OT grammar $G_{\gg}$ corresponding to a ranking $\gg$ maps an underlying form a to a candidate $(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}})$ which is preferred by the ranking $\gg$ to all other candidates $(\mathsf{a}, \mathsf{c}, \rho_{\mathsf{a},\mathsf{c}})$ which share that underlying form a.[12] Which conditions on the candidate and constraint sets ensure that the the OT grammars corresponding to any ranking is idempotent?

[11] The assumption that a grammar G maps an underlying form to a *single* candidate is not crucial: the analyses developed here extend to a framework where G maps an underlying form to a *set* of candidates, thus modeling phonological variation.

[12] As noted in footnote 11, I assume that grammars map an underlying form to a single candidate. This condition holds in OT provided the constraint set is sufficiently rich relative to the candidate set: for any two candidates $(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}})$ and $(\mathsf{a}, \mathsf{c}, \rho_{\mathsf{a},\mathsf{c}})$ which share the underlying form a, the constraint set contains a constraint C which prefers one of the two.

3.2.2 A condition on the candidate set: the no-breaking assumption

Let me start with candidate conditions for OT idempotency. Kubozono et al (2008) report that English *frog* is imported as [fu.róg.gu] into Japanese: the velar stop geminates (despite being voiced) because of a requirement on the placement of stress, captured here through a place-holder constraint STRESS. Assume an analysis of consonant gemination in terms of breaking of a single underlying consonant into two surface copies. The OT grammar (33) fails at idempotency: **b** = [ɡ] is phonotactically licit, as it is the surface realization of **a** = /ŋ/, which $\text{IDENT}_{[\text{nasal}]}$ prevents from geminating; yet **b** = /ɡ/ cannot surface faithfully, because $\text{IDENT}_{[\text{nasal}]}$ fails at protecting it against gemination.

(33)

/ŋ/	$\text{IDENT}_{[\text{nas}]}$	STRESS
☞ **a** = ŋ | **b** = ɡ	*	*
a = ŋ |\\ **c** = ɡɡ	**!	

/ɡ/	$\text{IDENT}_{[\text{nas}]}$	STRESS
b = ɡ | **b** = ɡ		*!
☞ **b** = ɡ |\\ **c** = ɡɡ		

This example shows that idempotency fails even in the simplest cases when correspondence relations break an underlying segment into multiple surface segments. The *no-breaking assumption* is thus necessary for idempotency.

3.2.3 A condition on the faithfulness constraint set: the FIC^{OT}_{comp}

Turning to constraint conditions for idempotency, the following Proposition 1 ensures idempotency when all the faithfulness constraints in the constraint set satisfy the implication (34). This implication will be referred to as the *faithfulness idempotency condition* (FIC^{OT}), as it only concerns the faithfulness constraints. No assumptions are made on the markedness constraints, on the nature of the faithfulness constraints (for instance, they are not required to be categorical), on the correspondence relations in the candidate set (for instance, they are not forbidden to break any underlying candidate). Proposition 1 is derived in Magri (to appear, section 3), by mimicking Tesar's (2013) analogous result for output-drivenness (see Proposition 8 below); see also Moreton and Smolensky (2002, section 3), Prince (2007), and Buccola (2013).

Proposition 1 *Assume that, for any two candidates $(\boldsymbol{a}, \boldsymbol{b}, \rho_{\boldsymbol{a},\boldsymbol{b}})$ and $(\boldsymbol{b}, \boldsymbol{c}, \rho_{\boldsymbol{b},\boldsymbol{c}})$, the candidate set also contains a candidate $(\boldsymbol{a}, \boldsymbol{c}, \rho_{\boldsymbol{a},\boldsymbol{c}})$ such that the following implication (34) holds for every faithfulness constraint F in the constraint set.*

(34) **If:** $F\big(\boldsymbol{b}, \boldsymbol{c}, \rho_{\boldsymbol{b},\boldsymbol{c}}\big) = 0$
Then: $F\big(\boldsymbol{a}, \boldsymbol{c}, \rho_{\boldsymbol{a},\boldsymbol{c}}\big) \leq F\big(\boldsymbol{a}, \boldsymbol{b}, \rho_{\boldsymbol{a},\boldsymbol{b}}\big)$

Then, the OT grammar corresponding to any ranking of the constraint set is idempotent, no matter what the markedness constraints look like. □

Proposition 1 makes no assumptions on the relationship between the correspondence relation $\rho_{\mathsf{a},\mathsf{c}}$ and the other two correspondence relations $\rho_{\mathsf{a},\mathsf{b}}$ and $\rho_{\mathsf{b},\mathsf{c}}$. A natural assumption is that $\rho_{\mathsf{a},\mathsf{c}}$ is the composition $\rho_{\mathsf{a},\mathsf{b}}\rho_{\mathsf{b},\mathsf{c}}$ of $\rho_{\mathsf{a},\mathsf{b}}$ and $\rho_{\mathsf{b},\mathsf{c}}$, whose existence is guaranteed by the transitivity axiom (6). The $\mathrm{FIC}^{\mathrm{OT}}$ (34) can thus be specialized as (35), which will be referred to as the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$.

(35) **If:** $F\big(\mathsf{b}, \mathsf{c}, \rho_{\mathsf{b},\mathsf{c}}\big) = 0$
Then: $F\big(\mathsf{a}, \mathsf{c}, \underbrace{\rho_{\mathsf{a},\mathsf{b}}\rho_{\mathsf{b},\mathsf{c}}}_{\rho_{\mathsf{a},\mathsf{c}}}\big) \leq F\big(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}}\big)$

The sufficient condition for idempotency provided by the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ is *tight*: any faithfulness constraint which fails the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ admits an elementary counterexample where idempotency fails (Magri to appear, section 7).

3.2.4 Which faithfulness constraints satisfy the FIC^{OT}_{comp}

Proposition 1 has established the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ (35) as a sufficient condition for idempotency. Which faithfulness constraints satisfy it? Magri (to appear, sections 4-6) addresses this question for a variety of faithfulness constraints within McCarthy and Prince's (1995) Correspondence Theory and its more recent extensions. The results are summarized in the following proposition.

Proposition 2 *Assume the candidate set (2) satisfies the transitivity axiom (6).*

(a) *The following constraints satisfy the FIC^{OT}_{comp} under no additional assumptions: segmental* Max, *featural* $\text{Max}_{[\pm\varphi]}$ *(Casali 1998),* Integrity.

(b) *The following constraints satisfy the FIC^{OT}_{comp} under the assumption that no correspondence relation in the candidate set breaks any underlying segment:* Ident_{φ} *(corresponding to a total feature φ), the local disjunction of any two identity faithfulness constraints (Downing 2000), segmental* Dep, *featural* $\text{Dep}_{[\mp\varphi]}$ *(Casali 1998),* Uniformity, Linearity, Max/DepLinearity *(Heinz 2005),* I-O/Adjacency *(Carpenter 2002).*

(c) *The following constraints fail at the FIC^{OT}_{comp} even when the correspondence relations in the candidate set are all one-to-one (no breaking nor coalescence):* Contiguity *and the local conjunction of any two conjoinable faithfulness constraints (Smolensky 1995).* □

The intuitive difference between the constraints listed in Proposition 2a (which satisfy the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ under no additional assumptions) and those listed in Proposition 2b (which instead require the no-breaking assumption) is that the former set of constraints only "count" over underlying segments (in the sense of Subsection 2.3.2) while the latter set of constraints (also) "count" over surface segments. Counterexample (33) repeated in (36) motivates the no-breaking assumption in Proposition 2b: $\text{Ident}_{[\text{nas}]}$ otherwise fails the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$.

(36) $\textsc{Id}_{[\text{nas}]}(\mathbf{a},\mathbf{b},\rho_{\mathbf{a},\mathbf{b}})=1$ $\qquad$ $\textsc{Id}_{[\text{nas}]}(\mathbf{b},\mathbf{c},\rho_{\mathbf{b},\mathbf{c}})=0$ $\qquad$ $\textsc{Id}_{[\text{nas}]}(\mathbf{a},\mathbf{c},\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})=2$

$\mathbf{a} =$ ŋ, $\mathbf{b} =$ g $\qquad$ $\mathbf{b} =$ g, $\mathbf{c} =$ gg $\qquad$ $\mathbf{a} =$ ŋ, $\mathbf{c} =$ gg

The no-breaking assumption of Proposition 2b is not too restrictive, as idempotency indeed fails in the presence of breaking, as seen in Subsection 3.2.2.

3.3 OT Idempotency and the triangle inequality

The condition for idempotency provided by the $\text{FIC}^{\text{OT}}_{\text{comp}}$ (35) looks like a technical condition without an intuitive interpretation. The rest of this section will derive an intuitive metric interpretation of this condition by investigating its relationship with the FTI_{comp} introduced in Section 2, repeated in (37).

(37) $F\big(\mathbf{a},\mathbf{c},\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}\big) \leq F\big(\mathbf{a},\mathbf{b},\rho_{\mathbf{a},\mathbf{b}}\big) + F\big(\mathbf{b},\mathbf{c},\rho_{\mathbf{b},\mathbf{c}}\big)$

3.3.1 The FTI_{comp} is stronger than the FIC^{OT}_{comp} in the general case

The FTI_{comp} entails the $\text{FIC}^{\text{OT}}_{\text{comp}}$ in the general case. In fact, assume that the antecedent of the $\text{FIC}^{\text{OT}}_{\text{comp}}$ (35) holds, namely that $F(\mathbf{b},\mathbf{c},\rho_{\mathbf{b},\mathbf{c}}) = 0$. In this case, the FTI_{comp} (37) coincides with the consequent of the $\text{FIC}^{\text{OT}}_{\text{comp}}$, which therefore holds true. The reverse entailment fails in the general case, as shown by the counterexample (38). The candidate $(\mathbf{a},\mathbf{b},\rho_{\mathbf{a},\mathbf{b}})$ in (38a) does not violate the faithfulness constraint Max: the two underlying consonants of the string $\mathbf{a}$ are coalesced into the single consonant of $\mathbf{b}$ but not deleted. The candidate $(\mathbf{b},\mathbf{c},\rho_{\mathbf{b},\mathbf{c}})$ in (38b) does violate Max once, because no correspondence is established between the underlying and surface consonants. The composition candidate $(\mathbf{a},\mathbf{c},\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$ in (38c) violates Max twice, as both underlying consonants lack a correspondent according to the composition correspondence relation $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$. Max thus fails at the FTI_{comp} (37): the left-hand side is equal to 2 and is thus larger than the right-hand side which is instead equal to $0+1$.

(38) a. $\textsc{Max}(\mathbf{a},\mathbf{b}) = 0$ $\qquad$ b. $\textsc{Max}(\mathbf{b},\mathbf{c}) = 1$ $\qquad$ c. $\textsc{Max}(\mathbf{a},\mathbf{c}) = 2$

$\mathbf{a} =$ CCV, $\mathbf{b} =$ CV $\qquad$ $\mathbf{b} =$ CV, $\mathbf{c} =$ CV $\qquad$ $\mathbf{a} =$ CCV, $\mathbf{c} =$ CV

Since Max satisfies the $\text{FIC}^{\text{OT}}_{\text{comp}}$ (under no assumptions on the correspondence relations) by Proposition 2a, the counterexample (38) shows that the FTI_{comp} is a stronger condition than the $\text{FIC}^{\text{OT}}_{\text{comp}}$ in the general case.

3.3.2 The FTI_{comp} and the FIC^{OT}_{comp} are equivalent for binary constraints

Which assumptions on the candidates and the faithfulness constraints suffice to take the edge off the FTI_{comp}, making it equivalent to the $\text{FIC}^{\text{OT}}_{\text{comp}}$? One of the properties of the counterexample (38) just used to show that the FTI_{comp} is stronger than the $\text{FIC}^{\text{OT}}_{\text{comp}}$ is that the composition candidate $(\mathsf{a}, \mathsf{c}, \rho_{\mathsf{a},\mathsf{b}}\rho_{\mathsf{b},\mathsf{c}})$ which appears on the left-hand side of the FTI_{comp} (37) incurs a large (namely, larger than 1) number of violations. To start, suppose instead that the quantity $F(\mathsf{a}, \mathsf{c}, \rho_{\mathsf{a},\mathsf{b}}\rho_{\mathsf{b},\mathsf{c}})$ which appears on the left-hand side of the FTI_{comp} is "small", namely equal to either 0 or 1 but not larger. Then, the $\text{FIC}^{\text{OT}}_{\text{comp}}$ entails the FTI_{comp}. In fact, if $F(\mathsf{b}, \mathsf{c}, \rho_{\mathsf{b},\mathsf{c}}) \geq 1$, then the right-hand side of the FTI_{comp} is already large enough to exceed the small left-hand side, ensuring that the inequality holds. If instead $F(\mathsf{b}, \mathsf{c}, \rho_{\mathsf{b},\mathsf{c}}) = 0$, the antecedent of the $\text{FIC}^{\text{OT}}_{\text{comp}}$ holds and the consequent of the $\text{FIC}^{\text{OT}}_{\text{comp}}$ then entails the FTI_{comp}. I conclude that the $\text{FIC}^{\text{OT}}_{\text{comp}}$ and the FTI_{comp} are equivalent when the faithfulness constraint F is *binary*, namely it assigns to any candidate either 0 or 1 violations.

3.3.3 The FTI_{comp} and the FIC^{OT}_{comp} are equivalent for categorical constraints

Most faithfulness constraints in the phonological literature are *not* binary.[13] The equivalence between the FTI_{comp} and the $\text{FIC}^{\text{OT}}_{\text{comp}}$ established in Subsection 3.3.2 for binary faithfulness constraints thus has a modest applicability. Yet, McCarthy (2003b) conjectures that all faithfulness constraints relevant for natural language phonology are categorical. This conjecture has been formalized and slightly strengthened in Section 2.3. Crucially, the equivalence between the FTI_{comp} and the $\text{FIC}^{\text{OT}}_{\text{comp}}$ extends from binary to categorical constraints. The intuitive idea is as follows. Consider a faithfulness constraint which satisfies the $\text{FIC}^{\text{OT}}_{\text{comp}}$. Assume that it is furthermore additive, so that the number of violations it assigns to a candidate is the sum of the numbers of violations it assigns to the sub-candidates. Assume finally that it is also categorical, so that these sub-candidates can be chosen in such a way that they are assigned either 0 or 1 violations. In other words, the faithfulness constraint is binary when restricted to the sub-candidates. By reasoning as in Subsection 3.3.2 for binary faithfulness constraints, the $\text{FIC}^{\text{OT}}_{\text{comp}}$ entails the FTI_{comp} when restricted to the sub-candidates. By summing over all sub-candidates through additivity, the FTI_{comp} for the original candidate finally follows. This intuitive reasoning is formalized in Appendix A.1 into a proof of the following Proposition 3, which is the main technical result of this paper. Since the triangle inequality characterizes the intuitive notion of distance, this proposition provides an intuitive interpretation of the sufficient condition for OT idempotency provided by the $\text{FIC}^{\text{OT}}_{\text{comp}}$ (35): the $\text{FIC}^{\text{OT}}_{\text{comp}}$ simply requires the faithfulness constraints to measure the phonological distance between underlying and surface forms in a sensible way, namely in compliance with the triangle inequality.

[13] McCarthy and Prince's (1995) definition of I-/O-Contiguity makes them binary.

Proposition 3 *Assume the candidate set (2) satisfies the transitivity axiom (6) and only contains one-to-one correspondence relations. Consider a faithfulness constraint F which is C-categorical; or I-categorical and O-monotone; or O-categorical and I-monotone. F satisfies the* FTI_{comp} *(37) if and only if it satisfies the* FIC^{OT}_{comp} *(35).* □

3.3.4 Some remarks on the assumptions of Proposition 3

Proposition 3 requires all correspondence relations in the candidate set to be one-to-one, namely rules out both coalescence and breaking (deletion and epenthesis are instead of course allowed).[14] This assumption is unavoidable. In fact, the counterexample (36) shows that the $\mathrm{FTI_{comp}}$ fails for core faithfulness constraints when correspondence relations can break an underlying segment into multiple surface segments. Furthermore, the counterexample (38) shows that it also fails when correspondence relations can coalesce multiple underlying segments into a single surface segment.

Proposition 3 also requires I- and O-categorical faithfulness constraints to satisfy O- and I-monotonicity while no monotonicity conditions are imposed on C-categorical faithfulness constraints. The proof provided in Appendix A.1 shows that this asymmetry between I/O-categoricity and C-categoricity stems from the asymmetry in the definitions of additivity discussed in Subsection 2.3.3: in the case of C-additivity (18), the sub-candidates are obtained by restricting only the correspondence relation; in the case of I-/O-additivity (23)/(29), the sub-candidates are obtained by restricting both the underlying/surface string and the correspondence relation.

The following counterexample shows that the additional monotonicity assumption made by Proposition 3 is indeed crucial. Consider the (unreasonable) faithfulness constraint F defined in (39) through the by now familiar two steps. It is indeed a faithfulness constraint, namely it assigns zero violations to the identity candidate, in compliance with condition (8).

(39) a. $F\big(\mathsf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}\big) = \begin{cases} 1 & \text{if the underlying segment } \mathsf{a} \text{ has no surface correspondents and the string } \mathbf{b} \text{ has length } 1 \\ 0 & \text{otherwise} \end{cases}$

b. $F\big(\mathbf{a},\, \mathbf{b},\, \rho_{\mathbf{a},\mathbf{b}}\big) = \sum_{\mathsf{a} \subseteq \mathbf{a}} F\big(\mathsf{a},\, \mathbf{b},\, \rho_{\mathbf{a},\mathbf{b}} \restriction_{(\mathsf{a},\mathbf{b})}\big)$

This constraint F is I-categorical (of order $\ell = 1$) by construction. Yet, the counterexample (40) shows that it is not O-monotone. In fact, F assigns one violation to the candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ in (40a), because the underlying coda /t/

[14] A close look at the proof of Proposition 3 shows that, if the faithfulness constraint F is I-categorical of order $\ell = 1$, the one-to-one assumption can be weakened to the assumption that no correspondence relation coalesces any two underlying segments into a single surface segment. If the faithfulness constraint F is O-categorical of order $\ell = 1$, the one-to-one assumption can be weakened to the assumption that no correspondence relation breaks any underlying segment into two surface segments. If the faithfulness constraint F is instead C-categorical, the one-to-one assumption cannot be weakened, not even in the case $\ell = 1$.

is deleted and the surface string $\mathbf{b}$ has length 1. O-monotonicity says that the number of violations cannot decrease if the surface string $\mathbf{b}$ is extended to $\widehat{\mathbf{b}}$ by adding surface segments. This requirement is shown to fail in (40b): the addition of the onset [p] increases the length of the surface string $\widehat{\mathbf{b}}$ to 2 and thus prevents F from assigning any violations to the candidate $(\mathbf{a}, \widehat{\mathbf{b}}, \rho_{\mathbf{a},\mathbf{b}})$

(40) a. $F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) = 1$

$$\begin{array}{rl} \mathbf{a} = & \mathsf{a\,t} \\ & \mathsf{|} \\ \mathbf{b} = & \mathsf{a} \end{array}$$

b. $F(\mathbf{a}, \widehat{\mathbf{b}}, \rho_{\mathbf{a},\mathbf{b}}) = 0$

$$\begin{array}{rl} \mathbf{a} = & \phantom{\mathsf{p}}\,\mathsf{a\,t} \\ & \phantom{\mathsf{p}}\,\mathsf{|} \\ \widehat{\mathbf{b}} = & \mathsf{p\,a} \end{array}$$

This faithfulness constraint F is easily shown to satisfy the $\text{FIC}^{\text{OT}}_{\text{comp}}$ (35) when correspondence relations are one-to-one.[15] Yet, F does not satisfy the FTI_{comp}, as shown by the counterexample (41). The candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ in (41a) does not violate F, because its surface string is longer than 1. The candidate $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ in (41b) violates F once, because it has a single deleted underlying segment. The right-hand side of the FTI_{comp} is thus equal to 0+1 and is smaller than the left-hand side, which is instead equal to 2 because the composition candidate $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$ in (41c) violates F twice.

(41) a. $F\big(\mathbf{a},\, \mathbf{b},\, \rho_{\mathbf{a},\mathbf{b}}\big) = 0$

$$\begin{array}{rl} \mathbf{a} = & \mathsf{a\,t\,t} \\ & \mathsf{|\,\,|} \\ \mathbf{b} = & \mathsf{a\,t} \end{array}$$

b. $F\big(\mathbf{b},\, \mathbf{c},\, \rho_{\mathbf{b},\mathbf{c}}\big) = 1$

$$\begin{array}{rl} \mathbf{b} = & \mathsf{a\,t} \\ & \mathsf{|} \\ \mathbf{c} = & \mathsf{a} \end{array}$$

c. $F\big(\mathbf{a},\, \mathbf{c},\, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}\big) = 2$

$$\begin{array}{rl} \mathbf{a} = & \mathsf{a\,t\,t} \\ & \mathsf{|} \\ \mathbf{c} = & \mathsf{a} \end{array}$$

In conclusion, the assumption made by Proposition 3 that I-categorical (and O-categorical) faithfulness constraints also be O-monotone (and I-monotone, respectively) is unavoidable in order for the $\text{FIC}^{\text{OT}}_{\text{comp}}$ to entail the FTI_{comp}.

3.3.5 Which faithfulness constraints satisfy the FTI_{comp}

I am now in a position to address the question raised at the end of Subsection 2.2: which faithfulness constraints satisfy the FTI_{comp} and thus measure the phonological distance between underlying and surface forms in compliance with the triangle inequality? In fact, Section 2.3 has shown that $\textsc{Ident}_{\varphi}$, I/O-$\textsc{Adjacency}$, and $\textsc{Linearity}$ are C-categorical; that $\textsc{Max}$, $\textsc{Max}_{[+\varphi]}$,

[15] In fact, assume that the quantity $F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$ in the left-hand side of the inequality in the consequent of the $\text{FIC}^{\text{OT}}_{\text{comp}}$ is larger than 0 (otherwise, the inequality trivially holds). This means that $\mathbf{c}$ has length 1. The antecedent $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) = 0$ of the $\text{FIC}^{\text{OT}}_{\text{comp}}$ thus requires every segment of $\mathbf{b}$ to have a correspondent in $\mathbf{c}$. Since correspondence relations are one-to-one, the string $\mathbf{b}$ must consist of a single segment which is put in correspondence by $\rho_{\mathbf{b},\mathbf{c}}$ with the single segment of $\mathbf{c}$. It then follows that every underlying segment a of $\mathbf{a}$ which violates F relative to the composition candidate $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$ also violates F relative to the candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$, thus establishing the inequality in the consequent of the $\text{FIC}^{\text{OT}}_{\text{comp}}$.

and MAXLINEARITY are I-categorical and O-monotone; finally, that DEP, $\text{DEP}_{[+\varphi]}$, and DEPLINEARITY are O-categorical and I-monotone.[16] The following Proposition 4 thus follows straightforwardly from the equivalence between the $\text{FIC}^{\text{OT}}_{\text{comp}}$ and the FTI_{comp} guaranteed by Proposition 3 together with the characterization of the faithfulness constraints which satisfy the $\text{FIC}^{\text{OT}}_{\text{comp}}$ summarized above as Proposition 2.

Proposition 4 *Assume the candidate set (2) satisfies the transitivity axiom (6) and only contains one-to-one correspondence relations. The following faithfulness constraints satisfy the FTI_{comp}: segmental* MAX *and* DEP*, featural* $\text{MAX}_{[\pm\varphi]}$ *and* $\text{DEP}_{[\pm\varphi]}$, IDENT_{φ} *(corresponding to a total feature φ), the local disjunction of any two identity constraints,* LINEARITY*,* MAXLINEARITY*,* DEPLINEARITY*, and* I/O-ADJACENCY*.*

On the other hand, CONTIGUITY and the local conjunction of any two conjoinable faithfulness constraints fail at the FTI_{comp}.

3.4 Summary

This section has shown that the sufficient condition for OT idempotency provided by the $\text{FIC}^{\text{OT}}_{\text{comp}}$ admits a metric interpretation: it effectively requires the faithfulness constraints to measure the phonological distance between underlying and surface representations in compliance with the metric triangle inequality. This interpretation holds under (a slightly stronger version of) McCarthy's (2003b) categoricity conjecture, formalized in Subsection 2.3.

4 Idempotency in Harmonic Grammar

This section extends the theory of idempotency from OT to HG. This extension will provide a more pristine view of the relationship between idempotency and the faithfulness triangle inequality, which does not rely on categoricity.

4.1 Candidate and constraint conditions for HG idempotency

4.1.1 Harmonic Grammar

As OT, also HG presupposes an underlying candidate set together with a constraint set which extracts the relevant phonological properties of the candidates. A *constraint weighting* $\boldsymbol{\theta}$ assigns a numerical weight $\theta_C \geq 0$ to each constraint C. The weighting $\boldsymbol{\theta}$ *prefers* a candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ to another candidate $(\mathbf{c}, \mathbf{d}, \rho_{\mathbf{c},\mathbf{d}})$ provided the weighted sum of the constraint violations of

[16] Analogous considerations hold for INTEGRITY (which is I-categorical and O-monotone) and UNIFORMITY (which is O-categorical and I-monotone). I ignore these two constraints here, because Proposition 4 requires correspondence relations to be one-to-one.

the former candidate is smaller than that of the latter, as in (42). The HG grammar $G_{\boldsymbol{\theta}}$ corresponding to a constraint weighting $\boldsymbol{\theta}$ maps an underlying form $\mathbf{a}$ to a candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ which is preferred by the weighting $\boldsymbol{\theta}$ to all other candidates $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$ which share that underlying form $\mathbf{a}$.

(42) $$\sum_C \theta_C C(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) < \sum_C \theta_C C(\mathbf{c}, \mathbf{d}, \rho_{\mathbf{c},\mathbf{d}})$$

Constraints are always interpreted as expressing penalties, never rewards. Hence, constraint weights need to be enforced to satisfy the non-negativity condition $\theta_C \geq 0$ in order for HG to avoid undesired typological predictions, whereby less marked structures are mapped to more marked ones (Legendre et al 2006; Keller 2000). This assumption that constraint weights are non-negative plays a crucial role in the rest of this section (in particular in the proof of Proposition 5 below provided in Appendix A.2).

4.1.2 A condition on the candidate set: the one-to-one assumption

The counterexample in Subsection 3.2.2 has shown that OT idempotency fails when correspondence relations can break an underlying segment into multiple surface segments. The same counterexample of course shows that the no-breaking assumption is needed for HG idempotency as well. The situation is different for coalescence of two underlying segments into a single surface segment. While coalescence does not threaten OT idempotency, it does hinder HG idempotency, as shown by the following counterexample. Before the suffix /-i/, Finnish long low vowels shorten (/aa/ → [a]), short low vowels raise (/a/ → [o]), short round vowels surface faithfully (/o/ → [o]) (Łubowicz 2011).

(43)			*sing nom*	*plural essive*	
	a.	/aa/ → [a]:	m**aa**	m**a**-i-na	‘earth’
			vap**aa**	vap**a**-i-na	‘free’
	b.	/a/ → [o]:	kiss**a**	kiss**o**-i-na	‘cat’
			vap**a**	vap**o**-i-na	‘fishing rod’
	c.	/o/ → [o]:	tal**o**	tal**o**-i-na	‘house’
			pelk**o**	pelk**o**-i-na	‘fear’

Let’s analyze vowel shortening as coalescence of two underlying vowels into a single surface vowel. The mapping /aa/ → [o] thus violates the faithfulness constraint $F = \text{IDENT}_{[\text{low}]}$ twice, while the mapping /a/ → [o] violates it only once, as represented in (44). Assume that the constraint set also contains a markedness constraint M which punishes [a]. The HG grammar corresponding to the weights $\theta_F = 2$ and $\theta_M = 3$ yields the chain shift (43) and thus fails at idempotency: the short low vowel [a] is phonotactically licit, as it is the surface realization of the long low vowel /aa/ which $\text{IDENT}_{[\text{low}]}$ prevents from raising; yet, /a/ does not surface faithfully, because $\text{IDENT}_{[\text{low}]}$ fails to protect it.

(44)

/aa/	$F = \text{IDENT}_{[\text{low}]}$	M
☞ [a]		$*$
[o]	$**$	

/a/	$F = \text{IDENT}_{[\text{low}]}$	M
[a]		$*$
☞ [o]	$*$	

In conclusion, HG idempotency fails even in the simplest cases when correspondence relations display breaking or coalescence, thus effectively requiring the correspondence relations in the candidate set to be one-to-one.

4.1.3 A condition on the faithfulness constraint set: the FIC^{HG}_{comp}

The following Proposition 5 provides the HG analogue of the OT Proposition 1. The proof is provided in Appendix A.2. Condition (45) will be referred to as the *HG faithfulness idempotency condition* (FIC^{HG}). No assumptions are made on the markedness constraints, on the nature of the faithfulness constraints (for instance, they are not required to be categorical), on the correspondence relations (for instance, they are not required to be one-to-one).

Proposition 5 *Assume that, for any two candidates* $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ *and* $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$*, the candidate set also contains a candidate* $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$ *such that the following condition (45) holds for any faithfulness constraint* F *in the constraint set.*

(45) *For every choice of the constant* $\xi \geq 0$*:*[17]
If: $F\big(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}\big) \leq \xi$
Then: $F\big(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}\big) \leq F\big(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}\big) + \xi$

Then, the HG grammar corresponding to any weighting of the constraint set is idempotent, no matter what the markedness constraints look like. □

The FIC^{HG} (45) can be specialized to the $\text{FIC}^{\text{HG}}_{\text{comp}}$ (46), by choosing as the candidate $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$ the composition $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$ of $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ and $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$, whose existence is guaranteed by the transitivity axiom (6).

(46) For every choice of the constant $\xi \geq 0$:
If: $F\big(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}\big) \leq \xi$
Then: $F\big(\mathbf{a}, \mathbf{c}, \underbrace{\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}}_{\rho_{\mathbf{a},\mathbf{c}}}\big) \leq F\big(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}\big) + \xi$

This sufficient condition for idempotency provided by the $\text{FIC}^{\text{HG}}_{\text{comp}}$ is *tight*: any faithfulness constraint which fails the $\text{FIC}^{\text{HG}}_{\text{comp}}$ can be shown to admit an elementary counterexample where HG idempotency fails.

[17] The implication (45) trivially holds also for $\xi < 0$, because the antecedent is always false in that case, due to the non-negativity of constraint violations.

4.2 HG idempotency and the triangular inequality

This section shows that the $\mathrm{FIC}^{\mathrm{HG}}_{\mathrm{comp}}$ is stronger than the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ so that the $\mathrm{FIC}^{\mathrm{HG}}_{\mathrm{comp}}$ is equivalent to the triangular inequality for any faithfulness constraint, while we have seen that the equivalence holds for the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ only under McCarthy's (strengthened) categoricity conjecture.

4.2.1 The FIC^{HG}_{comp} is stronger than the FIC^{OT}_{comp} in the general case

The sufficient condition for HG idempotency provided by the $\mathrm{FIC}^{\mathrm{HG}}_{\mathrm{comp}}$ (46) entails the sufficient condition for OT idempotency provided by the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$, repeated in (47) for ease of comparison. In fact, the latter is a special case of the former corresponding to the choice $\xi = 0$.

(47) **If:** $F\big(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}\big) = 0$
Then: $F\big(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}\big) \leq F\big(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}\big)$

The reverse implication fails in the general case, showing that the $\mathrm{FIC}^{\mathrm{HG}}_{\mathrm{comp}}$ is stronger than the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$. Here is a counterexample. The faithfulness constraint $\textsc{Ident}_{[\mathrm{low}]}$ satisfies the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ under the no-breaking assumption, by Proposition 2b. Yet, the counterexample (44) shows that it fails at the $\mathrm{FIC}^{\mathrm{HG}}_{\mathrm{comp}}$ (46) for the two corresponding candidates $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ and $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ in (48a) and (48b), their composition $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$ in (48c), and the constant $\xi = 1.5$.

(48) a. $\textsc{Idt}_{[\mathrm{low}]}(\mathbf{a}, \mathbf{b}) = 0$

a = a a
b = a

b. $\textsc{Idt}_{[\mathrm{low}]}(\mathbf{b}, \mathbf{c}) = 1$

b = a
c = o

c. $\textsc{Idt}_{[\mathrm{low}]}(\mathbf{a}, \mathbf{c}) = 2$

a = a a
c = o

The fact that the $\mathrm{FIC}^{\mathrm{HG}}_{\mathrm{comp}}$ is stronger than the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ in the general case is unsurprising, given that HG typologies are larger than OT typologies and thus harder to discipline to idempotency.

4.2.2 The FIC^{HG}_{comp} is equivalent to the FTI_{comp} in the general case

Subsection 2.2 has introduced the $\mathrm{FTI}_{\mathrm{comp}}$ repeated in (49) to formalize the condition that faithfulness constraints measure phonological distance in compliance with the triangle inequality. Suppose the antecedent of the $\mathrm{FIC}^{\mathrm{HG}}_{\mathrm{comp}}$ (46) holds. The term (*) in (49) is thus smaller than ξ and (49) entails the consequent of the $\mathrm{FIC}^{\mathrm{HG}}_{\mathrm{comp}}$. In other words, the $\mathrm{FTI}_{\mathrm{comp}}$ entails the $\mathrm{FIC}^{\mathrm{HG}}_{\mathrm{comp}}$.

(49) $$F\big(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}\big) \leq F\big(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}\big) + \underbrace{F\big(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}\big)}_{(*)}$$

The reverse entailment holds as well, because the antecedent of the $\mathrm{FIC}^{\mathrm{HG}}_{\mathrm{comp}}$ (46) trivially holds with the position $\xi = F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ and its consequent is equivalent to the $\mathrm{FTI}_{\mathrm{comp}}$ (49) in this case.

In conclusion, we have obtained the equivalence between the $\mathrm{FIC}^{\mathrm{HG}}_{\mathrm{comp}}$ and the $\mathrm{FTI}_{\mathrm{comp}}$ stated by the following Proposition 6. Since the triangle inequality characterizes the intuitive notion of distance, this proposition provides an intuitive interpretation of the sufficient condition for HG idempotency provided by the $\mathrm{FIC}^{\mathrm{HG}}_{\mathrm{comp}}$ (46): the $\mathrm{FIC}^{\mathrm{HG}}_{\mathrm{comp}}$ simply requires the faithfulness constraints to measure the phonological distance between underlying and surface forms in a sensible way, namely in compliance with the triangle inequality.

Proposition 6 *The FIC^{HG}_{comp} (46) and the FTI_{comp} repeated in (49) are equivalent: a faithfulness constraint satisfies the former relative to two candidates $(\boldsymbol{a}, \boldsymbol{b}, \rho_{\boldsymbol{a},\boldsymbol{b}})$ and $(\boldsymbol{b}, \boldsymbol{c}, \rho_{\boldsymbol{b},\boldsymbol{c}})$ if and only if it satisfies the latter.*

This Proposition 6 is analogous to Proposition 3 in Section 3, which established an equivalence between the $\mathrm{FTI}_{\mathrm{comp}}$ and the sufficient condition for OT idempotency provided by the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$. The crucial difference is that the latter proposition for OT required correspondence relations to be one-to-one and the faithfulness constraints to be categorical and monotone. Proposition 6 for HG instead does not make any assumption neither on the correspondence relations nor on the faithfulness constraints. The fact that the equivalence holds under weaker conditions in HG than in OT reflects the fact that the $\mathrm{FIC}^{\mathrm{HG}}_{\mathrm{comp}}$ is stronger than the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ in the general case, as noted in Subsection 4.2.1.

4.2.3 The FIC^{HG}_{comp} and the FIC^{OT}_{comp} are equivalent for categorical constraints

Assume that all correspondence relations in the candidate set are one-to-one, as otherwise HG idempotency fails (Subsection 4.1.2) and the $\mathrm{FIC}^{\mathrm{HG}}_{\mathrm{comp}}$ must fail as well. Assume furthermore that each faithfulness constraint in the constraint set satisfies McCarthy's (strengthened) categoricity conjecture, formalized in Subsection 2.3. The $\mathrm{FIC}^{\mathrm{HG}}_{\mathrm{comp}}$ and the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ are then equivalent, as they are both equivalent to the $\mathrm{FTI}_{\mathrm{comp}}$ by Propositions 3 and 6.

Proposition 7 *Assume that the correspondence relations in the candidate set are all one-to-one. Consider a faithfulness constraint F which is C-categorical; or I-categorical and O-monotone; or O-categorical and I-monotone. F satisfies the FIC^{HG}_{comp} (46) if and only if it satisfies the FIC^{OT}_{comp} repeated in (47).* □

Since an HG typology can be larger than an OT typology, we expect stronger conditions to be needed to discipline all the grammars in the former to satisfy idempotency. Indeed, candidate conditions for HG idempotency (no breaking and no coalescence) are stronger than candidate conditions for OT idempotency (no breaking only), as shown in Subsection 4.1.2. The constraint conditions instead turn out to be equivalent on the background of McCarthy's (strengthened) conjecture. The constraints listed by Proposition 2 in Section 3 as satisfying the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ thus also satisfy the $\mathrm{FIC}^{\mathrm{HG}}_{\mathrm{comp}}$ (under the further assumption that all correspondence relations are one-to-one).

4.3 Summary

This section has completed the theory of idempotency within the OT and HG implementations of constraint-based phonology. Idempotency has been shown to follow from conditions on the faithfulness constraints, namely the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ and the $\mathrm{FIC}^{\mathrm{HG}}_{\mathrm{comp}}$. These conditions are equivalent for faithfulness constraints which satisfy McCarthy's (strengthened) categoricity conjecture, because both conditions can be interpreted as requiring the faithfulness constraints to measure phonological distances in compliance with the triangle inequality, formalized as the $\mathrm{FTI}_{\mathrm{comp}}$. The rest of the paper turns to output-drivenness.

5 Output-drivenness in Optimality Theory

This section shows that the metric triangle inequality plays a crucial role also in the theory of Tesar's output-drivenness. To highlight this role, I introduce the notion of similarity (that output-drivenness is predicated on) axiomatically through a condition on the faithfulness constraints, rather than concretely in terms of strings and correspondence relations, as Tesar does.

5.1 Output-drivenness and its relationship of idempotency

This Subsection reviews Tesar's notion of output-drivennessand compares it to the notion of idempotency considered in the first half of the paper.

5.1.1 Tesar's notion of output-drivenness

Consider two candidates $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ and $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ which share a surface form $\mathbf{d}$. Suppose that the underlying form $\mathbf{b}$ is more similar to $\mathbf{d}$ than the other underlying form $\mathbf{a}$ is. In other words, that the candidate $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ has more *internal similarity* than the candidate $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$. Tesar formalizes this assumption through the condition (50), where $\leq_{\mathrm{sim}}$ is a *similarity order*, namely an ordering relation among candidates (or, more precisely, among candidates which share the surface form) based on their internal similarity. Subsections 5.3-5.4 will address the issue of the proper definition of this similarity order.

(50) $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\mathrm{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$

Suppose that a phonological grammar G maps the less similar underlying form $\mathbf{a}$ to the surface form $\mathbf{d}$, namely that $G(\mathbf{a}) = (\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$. This means that $\mathbf{d}$ is phonotactically licit and that $\mathbf{a}$ is not too dissimilar from $\mathbf{d}$. Since the phonotactic status of $\mathbf{d}$ does not depend on the underlying form and furthermore $\mathbf{b}$ is even more similar to $\mathbf{d}$, the grammar G should map also the more similar underlying form $\mathbf{b}$ to that same surface form $\mathbf{d}$, namely $G(\mathbf{b}) = (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$. A grammar which abides by this logics is called *output-driven*.

Definition 2 (Tesar's Output-drivenness) A grammar G is *output-driven* relative to a similarity order $\leq_{\text{sim}}$ provided the following implication holds

(51)	**If:**	$G(\mathsf{a}) = (\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}})$	*(candidate with less similarity)*
	Then:	$G(\mathsf{b}) = (\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}})$	*(candidate with more similarity)*

for any two candidates $(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}})$ and $(\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}})$ which share the surface form d and satisfy the condition $(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}}) \leq_{\text{sim}} (\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}})$. □

5.1.2 Output-drivenness entails idempotency

Intuitively, any string b is more similar to itself than to any other string a. In other words, identity candidates have the greatest internal similarity. A similarity order $\leq_{\text{sim}}$ thus needs to satisfy condition (52) for any candidate $(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}})$ and the corresponding identity candidate $(\mathsf{b}, \mathsf{b}, \mathbb{I}_{\mathsf{b},\mathsf{b}})$, whose existence is guaranteed by the reflexivity axiom (4).

(52) $(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}}) \leq_{\text{sim}} (\mathsf{b}, \mathsf{b}, \mathbb{I}_{\mathsf{b},\mathsf{b}})$

Whenever condition (52) holds, output-drivenness entails idempotency. In fact, output-drivenness requires the implication (51) to hold for any two candidates $(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}})$ and $(\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}})$ such that the former has less internal similarity than the latter. Condition (52) ensures that is indeed the case when the two strings b and d coincide and $\rho_{\mathsf{b},\mathsf{d}}$ is the identity on the string $\mathsf{b} = \mathsf{d}$). In this case, the implication (51) in the definition of output-drivenness specializes to (53), which is in turn the implication (32) in the definition of idempotency.

(53)	**If:**	$G(\mathsf{a}) = (\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}})$
	Then:	$G(\mathsf{b}) = (\mathsf{b}, \mathsf{b}, \mathbb{I}_{\mathsf{b},\mathsf{b}})$

In conclusion, the definition of idempotency coincides with the definition of output-drivenness in the special case where $\mathsf{b} = \mathsf{d}$ and $\rho_{\mathsf{b},\mathsf{d}}$ is the identity.

5.1.3 Output-drivenness is stronger than idempotency: saltations

Although output-drivenness entails idempotency, the reverse entailment fails: output-drivenness is a stronger condition than idempotency. As a counterexample, consider the *derived environment effect* or *saltation* (Łubowicz 2002, White 2013) in (54) from the Campidanian dialect of Sardinia: voiceless stops are lenited to voiced fricatives in post-vocalic position, while voiced stops are faithfully realized (Bolognesi 1998, via White 2013).

(54)	*isolated form*	*post-vocalic form*	
/p, t, k/ → [β, ð, ɣ]:	[**p**iʃ:i]	[bel:u **β**iʃ:i]	'(nice) fish'
	[**t**rintaduzu]	[s:u **ð**rintaduzu]	'(the) thirty-two'
	[**k**uat:ru]	[dɛ **ɣ**uat:ru]	'(of) four'
/b, d, g/ → [b, d, g]:	[**b**ĩu]	[s:u **b**ĩu]	'(the) wine'
	[**d**ominiɣu]	[don:ja **d**ominiɣu]	'(every) Sunday'
	[**g**ɔma]	[dɛ **g**ɔma]	'(of) rubber'

A reasonable definition of the similarity order $\leq_{\text{sim}}$ (see below footnote 22) should guarantee that the candidate $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ in (55) has less internal similarity than the candidate $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$, since the former involves a disparity for both voicing and continuancy, while the latter only for continuancy.

(55)
$$\begin{array}{ll} \mathbf{a} = /\mathsf{p}/ & \qquad \mathbf{b} = /\mathsf{b}/ \\ \qquad | & \qquad\qquad | \\ \mathbf{d} = [\beta] & \qquad \mathbf{d} = [\beta] \end{array}$$

Since $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$, output-drivenness would require the more similar underlying form $\mathbf{b} = /\mathsf{b}/$ to be mapped to the surface form $\mathbf{d} = [\beta]$ when the less similar underlying form $\mathbf{a} = /\mathsf{p}/$ is mapped to $\mathbf{d} = [\beta]$. The saltation pattern (54) thus fails at output-drivenness, despite being idempotent.

5.2 Output-driveness in OT

Which conditions on the candidate and constraint sets ensure that the OT grammars corresponding to any ranking is output-driven? This Subsection reviews Tesar's theory of output-drivenness in OT.

5.2.1 A condition on the candidate set: the one-to-one assumption

Let me start with candidate conditions. As seen in Subsection 3.2.2, OT idempotency fails when correspondence relations can break an underlying segment into multiple surface segments. Since output-drivenness entails idempotency, the no-breaking assumption is needed for output-drivenness as well. The situation is different for coalescence of two underlying segments into a single surface segment. While coalescence does not threaten OT idempotency, it does hinder output-drivenness, as shown by the following counterexample. A reasonable definition of the similarity order $\leq_{\text{sim}}$ (see below footnote 22) should guarantee that the candidate $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ in (56), with coalescence of the underlying complex coda, has less internal similarity than the candidate $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$.

(56)
$$\begin{array}{ll} \mathbf{a} = \mathsf{t\,e\,t\,d} & \qquad \mathbf{b} = \mathsf{t\,a\,d} \\ \quad\ \ |\ |\ \backslash\!/ & \qquad\quad\ \ |\ |\ | \\ \mathbf{d} = \mathsf{t\,a\,t} & \qquad \mathbf{d} = \mathsf{t\,a\,t} \end{array}$$

Although $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{c}})$, the constraint $M = {*}\textsc{Voice}$ in the grammar (57) maps the underlying form $\mathbf{a}$ to the candidate $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ with less internal similarity; yet, it fails at mapping the underlying form $\mathbf{b}$ to the candidate $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ with more internal similarity, violating output-drivenness.

(57)

/tetd/	$\textsc{Ident}_{[\text{vce}]}$	M
☞ **a** = tetd | | \/ **d** = tat	*	
a = tetd | | \/ **c** = tad	*	*!

/tad/	$\textsc{Ident}_{[\text{vce}]}$	M
b = tad | | | **d** = tat	*!	
☞ **b** = tad | | | **c** = tad		*

In conclusion, output-drivenness fails even in the simplest cases when correspondence relations display breaking or coalescence and thus are not one-to-one.

5.2.2 A condition on the faithfulness constraint set: the FODC^{OT}

The following Proposition 8 ensures output-drivenness when all the faithfulness constraints satisfy the two implications (58). These implications are thus jointly referred to as the OT *faithfulness output-drivenness condition* (FODC$^{\mathrm{OT}}$). No assumptions are made on the markedness constraints, on the nature of the faithfulness constraints (for instance, they are not required to be categorical), on the correspondence relations in the candidate set (for instance, they are not required to be one-to-one), or on the similarity order (which is left completely unspecified). Proposition 8 was derived in Tesar (2013, chapter 3). It is analogous to Proposition 1 in Subsection 3.2.3, which was indeed derived in Magri (to appear) by mimicking Tesar's reasoning.

Proposition 8 *Assume that, for any two candidates* $(\boldsymbol{a}, \boldsymbol{d}, \rho_{\boldsymbol{a},\boldsymbol{d}})$ *and* $(\boldsymbol{b}, \boldsymbol{d}, \rho_{\boldsymbol{b},\boldsymbol{d}})$ *such that* $(\boldsymbol{a}, \boldsymbol{d}, \rho_{\boldsymbol{a},\boldsymbol{d}}) \leq_{sim} (\boldsymbol{b}, \boldsymbol{d}, \rho_{\boldsymbol{b},\boldsymbol{d}})$*, for any other candidate* $(\boldsymbol{b}, \boldsymbol{c}, \rho_{\boldsymbol{b},\boldsymbol{c}})$ *different from* $(\boldsymbol{b}, \boldsymbol{d}, \rho_{\boldsymbol{b},\boldsymbol{d}})$*, the candidate set also contains a candidate* $(\boldsymbol{a}, \boldsymbol{c}, \rho_{\boldsymbol{a},\boldsymbol{c}})$ *different from* $(\boldsymbol{a}, \boldsymbol{d}, \rho_{\boldsymbol{a},\boldsymbol{d}})$ *such that the two following implications (58) hold for every faithfulness constraint F in the constraint set.*

(58) *a.* **If:** $F(\boldsymbol{a}, \boldsymbol{d}, \rho_{\boldsymbol{a},\boldsymbol{d}}) < F(\boldsymbol{a}, \boldsymbol{c}, \rho_{\boldsymbol{a},\boldsymbol{c}})$
Then: $F(\boldsymbol{b}, \boldsymbol{d}, \rho_{\boldsymbol{b},\boldsymbol{d}}) < F(\boldsymbol{b}, \boldsymbol{c}, \rho_{\boldsymbol{b},\boldsymbol{c}})$

b. **If:** $F(\boldsymbol{b}, \boldsymbol{c}, \rho_{\boldsymbol{b},\boldsymbol{c}}) < F(\boldsymbol{b}, \boldsymbol{d}, \rho_{\boldsymbol{b},\boldsymbol{d}})$
Then: $F(\boldsymbol{a}, \boldsymbol{c}, \rho_{\boldsymbol{a},\boldsymbol{c}}) < F(\boldsymbol{a}, \boldsymbol{d}, \rho_{\boldsymbol{a},\boldsymbol{d}})$

Then, the OT grammar corresponding to any ranking of the constraint set is output-driven relative to the similarity order $\leq_{sim}$. □

5.2.3 The FODC^{OT} entails the FIC^{OT}

Subsection 5.1.2 has shown that output-drivenness entails idempotency because the definition of idempotency coincides with the definition of output-drivenness in the special case of maximal similarity where the two strings $\mathbf{b}$ and $\mathbf{d}$ coincide. This entailment carries over to the two sufficient conditions for output-drivenness and idempotency: the FODC$^{\mathrm{OT}}$ entails the FIC$^{\mathrm{OT}}$. In fact, condition (52) on the similarity order ensures that the FODC$^{\mathrm{OT}}$ holds in the special case where the two strings $\mathbf{b}$ and $\mathbf{d}$ coincide and the correspondence relation $\rho_{\mathbf{b},\mathbf{d}}$ is the identity over the string $\mathbf{b} = \mathbf{d}$. In this special case, the contrapositive of the first FODC$^{\mathrm{OT}}$ implication (58a) specializes to (59).[18]

[18] The second FODC$^{\mathrm{OT}}$ implication (58b) is trivially satisfied in the special case where $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) = (\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$, because its antecedent $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) < F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ becomes $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) < 0$, which contradicts the non-negativity of constraint violations.

(59) **If:** $F(\mathsf{b}, \mathsf{c}, \rho_{\mathsf{b},\mathsf{c}}) \leq F(\mathsf{b}, \mathsf{b}, \mathbb{I}_{\mathsf{b},\mathsf{b}})$
Then: $F(\mathsf{a}, \mathsf{c}, \rho_{\mathsf{a},\mathsf{c}}) \leq F(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}})$

By (8), faithfulness constraints assign no violations to identity candidates. Since violation numbers are non-negative, the antecedent of (59) is equivalent to the condition $F(\mathsf{b}, \mathsf{c}, \rho_{\mathsf{b},\mathsf{c}}) = 0$. The (contrapositive of) the first FODC$^{\text{OT}}$ implication in (59) thus coincides with the FIC$^{\text{OT}}$ (34).

5.2.4 The FODC^{OT} is stronger than the FIC^{OT}

Subsection 5.1.3 has shown that output-drivenness is stronger than idempotency, because output-drivenness also excludes idempotent phonological patterns such as saltations. This relationship carries over to the two sufficient conditions for output-drivenness and idempotency: the FODC$^{\text{OT}}$ is stronger than the FIC$^{\text{OT}}$ (34), as shown by the following counterexample. The faithfulness constraint $F = \text{IDENT}_{[\text{voice}]} \vee \text{IDENT}_{[\text{cont}]}$ is the disjunction of the two identity constraints $\text{IDENT}_{[\text{voice}]}$ and $\text{IDENT}_{[\text{cont}]}$ for voicing and continuancy (Downing 1998, 2000). Proposition 2b ensures that F satisfies the FIC$^{\text{OT}}$ (provided the correspondence relations do not break any underlying segment). Yet, F fails at the FODC$^{\text{OT}}$. In fact, consider again the candidates $(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}})$ and $(\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}})$ in (55). As noted above, $(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}}) \leq_{\text{sim}} (\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}})$. Consider the position $\mathsf{b} = \mathsf{c}$. The antecedent of the second FODC$^{\text{OT}}$ implication (58b) holds, as shown in (60a). Yet, its consequent fails, as shown in (60b).

(60) a. $F(/\mathsf{b}/, [\mathsf{b}]) < F(/\mathsf{b}/, [\beta])$ namely $F(\mathsf{b}, \mathsf{c}) < F(\mathsf{b}, \mathsf{d})$
b. $F(/\mathsf{p}/, [\mathsf{b}]) = F(/\mathsf{p}/, [\beta])$ namely $F(\mathsf{a}, \mathsf{c}) = F(\mathsf{a}, \mathsf{d})$

5.2.5 Only a sufficient condition?

Consider an arbitrary candidate set, an arbitrary constraint set, and an arbitrary constraint ranking. Subsection 5.2.2 has established the FODC$^{\text{OT}}$ as a *sufficient* condition for the output-drivenness of the corresponding OT grammar. This statement contains three universal quantifications: over candidate sets, over constraint sets, and over rankings. At this level of generality, the FODC$^{\text{OT}}$ is not only a sufficient but also a *necessary* condition for output-drivenness. Let me illustrate this point with the faithfulness constraint $F = \text{IDENT}_{[\text{voice}]} \vee \text{IDENT}_{[\text{cont}]}$ which Subsection 5.2.4 has just shown to fail at the FODC$^{\text{OT}}$. Tableaux (61) show indeed that it can be straightforwardly used to derive the non-output-driven saltation in (54), whereby /p/ is mapped all the way to [β] while the closer /b/ is faithfully mapped to itself.

(61) a.

/p/	*[p]	F	*[p, b]
[p]	*		*
[b]		*	*
☞ [β]		*	

b.

/b/	*[p]	F	*[p, b]
[p]	*	*	*
☞ [b]			*
[β]		*	

The account in (61) has the following formal structure. The conjoined constraint F effectively requires perfect identity in both voicing and continuancy. Markedness constraints are split into those above F (here, a single constraint against voiceless labials; Flack 2007) and those below it (here, a single constraint which penalizes stops and thus favors spirantization). If the markedness constraints above F can be satisfied with perfect identity so that F is not violated, those constraints get to determine the winner. Otherwise, it is the markedness constraints below F which determine the winner.

5.3 Tesar's definition of the similarity order

To make further progress in the theory of output-drivenness, we need to make assumptions on the similarity order $\leq_{\text{sim}}$ that output-drivenness is predicated on. This Subsection reconstructs Tesar's definition of the similarity order, setting the ground for the axiomatization in the next Subsection.

5.3.1 An initial attempt

Tesar assumes a faithfulness constraint set which consists of the constraints MAX and DEP together with the set of featural IDENT constraints relative to a set Φ of (total) phonological features (see Magri 2016b for discussion of output-drivenness with partial phonological features). Within this framework, Definition 3 provides an intuitive definition of the similarity order, which captures the intuition that the less similar candidate $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ "makes up" for any underlying/surface disparity of the more similar candidate $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$.

Definition 3 For any candidates $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ and $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ sharing a surface form $\mathbf{d}$, let $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}}^{\Phi,\text{weak}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ provided:

Identity clause: for every feature φ in Φ, for every segment d of $\mathbf{d}$, if there exists a segment b of $\mathbf{b}$ such that $(\mathsf{b}, \mathsf{d}) \in \rho_{\mathbf{b},\mathbf{d}}$ and $\varphi(\mathsf{b}) \neq \varphi(\mathsf{d})$, then there exists a segment a of $\mathbf{a}$ such that $(\mathsf{a}, \mathsf{d}) \in \rho_{\mathbf{a},\mathbf{d}}$ and $\varphi(\mathsf{a}) \neq \varphi(\mathsf{d})$.

Deletion clause: there exists an injection from the segments of $\mathbf{b}$ deleted relative to $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ to the segments of $\mathbf{a}$ deleted relative to $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$.[19]

Epenthesis clause: every segment of $\mathbf{d}$ which is epenthetic relative to $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ is also epenthetic relative to $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$. $\square$

Let me illustrate Definition 3 with a couple of examples which will be relevant for what follows. The two candidates in (62a) satisfy the similarity inequality $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}}^{\Phi,\text{weak}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$. In fact, the epenthesis clause trivially holds because neither candidate displays epenthesis. The identity clause holds relative to the feature set $\Phi = \{[\text{high}], [\text{voice}]\}$, because the more similar candidate $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ displays no feature mismatches. Finally, the deletion clause

[19] This means that there exists a mapping from the segments of $\mathbf{b}$ deleted relative to $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ to the segments of $\mathbf{a}$ deleted relative to $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ such that two different deleted segments of $\mathbf{b}$ correspond to two different deleted segments of $\mathbf{a}$. In other words, $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ has at least as many deleted segments as $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$.

holds because both candidates feature exactly one deleted coda. Note crucially that the two deleted underlying codas differ from each other in voicing.

(62) a. $\mathbf{a}$ = r i d $\mathbf{b}$ = r e t   b. $\mathbf{a}$ = p a $\mathbf{b}$ = k i
  | | | | | | | |
$\mathbf{d}$ = r e $\mathbf{d}$ = r e $\mathbf{d}$ = t i $\mathbf{d}$ = t i

As another example, the two candidates in (62b) satisfy the similarity inequality $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}}^{\Phi,\text{weak}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$. In fact, the epenthesis and deletion clauses trivially hold, because neither candidate features epenthesis or deletion. Consider the feature set $\Phi = \{[\text{high}], [\text{place}]\}$ and assume that [place] is a three-valued feature. The identity clause holds because the underlying onset of both candidates has a surface correspondent which mismatches in place but only the underlying vowel of the less similar candidate has a surface correspondent which mismatches in height. Note crucially that the two underlying onsets differ from each other in place of articulation.

5.3.2 Towards a stronger similarity order

Unfortunately, output-drivenness fails even in the simplest cases relative to the similarity order provided by Definition 3. To start, consider the two candidates $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ and $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ in (62a). Suppose that, because of a higher ranked markedness constraint against voiced codas (omitted here), the OT grammar (63a) maps the underlying form $\mathbf{a}$ to the surface form $\mathbf{d}$ (which deletes the voiced coda) rather than to $\mathbf{c}$ (which instead devoices it). Since $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$, output-drivenness requires that grammar to map also the underlying form $\mathbf{b}$ to $\mathbf{d}$, contrary (63b).[20]

(63) a.

/rid/	ID$_{[\text{voice}]}$	MAX
$\mathbf{a}$ = rid | | | $\mathbf{c}$ = ret	*!	
☞ $\mathbf{a}$ = rid | | $\mathbf{d}$ = re		*

b.

/ret/	ID$_{[\text{voice}]}$	MAX
☞ $\mathbf{b}$ = ret | | | $\mathbf{c}$ = ret		
$\mathbf{b}$ = ret | | $\mathbf{d}$ = re		*

As another counterexample, consider the two candidates $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ and $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ in (62b). The OT grammar (64a) maps the underlying form $\mathbf{a}$ to the surface form $\mathbf{d}$. Since $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$, output-drivenness requires that grammar to map also the underlying form $\mathbf{b}$ to $\mathbf{d}$, contrary to (63b).

[20] It is an interesting exercise to check that this failure of output-drivenness comes, as expected, with a failure of the FODC$^{\text{OT}}$. Indeed, the first FODC$^{\text{OT}}$ (58a) fails for $F = \text{IDENT}_{[\text{voice}]}$: $\text{ID}(\mathbf{a}, \mathbf{d}) = 0$ and $\text{ID}(\mathbf{a}, \mathbf{c}) = 1$, so that the antecedent of the first FODC$^{\text{OT}}$ (58a) holds; yet $\text{ID}(\mathbf{b}, \mathbf{d}) = \text{ID}(\mathbf{b}, \mathbf{c}) = 0$, so that its consequent fails. Other choices of the correspondence relation $\rho_{\mathbf{a},\mathbf{c}}$ yields analogous failures. For instance, suppose that $\rho_{\mathbf{a},\mathbf{c}}$ establishes no correspondence between the codas of $\mathbf{a}$ and $\mathbf{c}$. In this case, the second FODC$^{\text{OT}}$ (58b) fails for $F = \text{MAX}$: $\text{MAX}(\mathbf{b}, \mathbf{c}) = 0$ and $\text{MAX}(\mathbf{b}, \mathbf{d}) = 1$, so that the antecedent of the second FODC$^{\text{OT}}$ (58b) holds; yet $\text{MAX}(\mathbf{a}, \mathbf{c}) = \text{MAX}(\mathbf{a}, \mathbf{d}) = 1$, so that its consequent fails.

(64) a.

/pa/	$\text{ID}_{[\text{place}]}$	*[k]
$\mathbf{a}$ = pa | | $\mathbf{b}$ = ki	$*$	$*!$
☞ $\mathbf{a}$ = pa | | $\mathbf{d}$ = ti	$*$	

b.

/ki/	$\text{ID}_{[\text{place}]}$	*[k]
☞ $\mathbf{b}$ = ki | | $\mathbf{b}$ = ki		$*$
$\mathbf{b}$ = ki | | $\mathbf{d}$ = ti	$*!$	

These counterexamples show that Definition 3 is too weak: it yields a similarity order which holds between too many pairs of candidates, making output-drivenness too hard to satisfy. We thus need a stronger definition of the similarity order, which is satisfied by less pairs of candidates and in particular is not satisfied by pairs of candidates such as those in (62), which we have just seen to be problematic for output-drivenness.

5.3.3 Tesar's similarity order

Indeed, the similarity order considered by Tesar is not the one provided by Definition 3 above but the one provided by the following Definition 4. The two definitions differ for the material underlined.

Definition 4 (Tesar's similarity order) For any candidates $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ and $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ sharing a surface form $\mathbf{d}$, let $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq^{\Phi}_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ provided:

Identity clause: for every feature φ in Φ, for every segment d of $\mathbf{d}$, if there exists a segment b of $\mathbf{b}$ such that $(\mathsf{b}, \mathsf{d}) \in \rho_{\mathbf{b},\mathbf{d}}$ and $\varphi(\mathsf{b}) \neq \varphi(\mathsf{d})$, there exists a segment a of $\mathbf{a}$ such that $(\mathsf{a}, \mathsf{d}) \in \rho_{\mathbf{a},\mathbf{d}}$ and $\varphi(\mathsf{a}) = \varphi(\mathsf{b})$.

Deletion clause: there exists an injection from the segments of $\mathbf{b}$ deleted relative to $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ to the segments of $\mathbf{a}$ deleted relative to $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ such that any two deleted segments corresponding through the injection agree on the value of every feature in Φ.

Epenthesis clause: every segment of $\mathbf{d}$ which is epenthetic relative to $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ is also epenthetic relative to $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$. □

The deletion clause of the new Definition 4 is stronger than the corresponding clause in the original Definition 3 because of the additional boxed condition which requires two deleted segments (corresponding through the injection) to match in feature values. To illustrate, consider the two candidates in (62a). The two deleted underlying codas differ relative to the feature [voice]. The additional underlined condition of the deletion clause therefore fails. The two candidates in (62a) thus fail the similarity inequality $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq^{\Phi}_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ relative to the similarity order $\leq^{\Phi}_{\text{sim}}$ provided by the revised Definition 4. The problem with the failure of output-drivenness in (63) is therefore circumvented.

The identity clause of the new Definition 4 has the underlined condition $\varphi(\mathsf{a}) = \varphi(\mathsf{b})$ instead of the condition $\varphi(\mathsf{a}) \neq \varphi(\mathsf{d})$ of the original Definition 3. Because of the assumption $\varphi(\mathsf{b}) \neq \varphi(\mathsf{d})$, the condition $\varphi(\mathsf{a}) = \varphi(\mathsf{b})$ of the new Definition 4 entails the condition $\varphi(\mathsf{a}) \neq \varphi(\mathsf{d})$ of the original Definition 3. The reverse entailment fails whenever the feature φ is has more than two values. To illustrate, consider the two candidates in (62b). The underlying onset /k/ of the underlying form **b** differs from its surface correspondent [t] relative to the feature [place]. The underlying onset /p/ of **a** also differs from its surface correspondent [t]. The identity condition of the original Definition 3 thus holds. But the identity condition of the new Definition 4 fails, because the two underlying onsets /p/ and /k/ differ relative to the three-valued feature [place]. The two candidates in (62b) thus fail the similarity inequality $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq^{\Phi}_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ relative to the similarity order $\leq^{\Phi}_{\text{sim}}$ provided by the revised Definition 4. The problem with the failure of output-drivenness in (64) is therefore circumvented.

5.4 Axiomatizing the similarity order

Tesar's Definition 4 describes the similarity order *concretely*, in terms of strings and correspondence relations. Within constraint-based phonology, it is natural to assess similarity through the faithfulness constraints. This Subsection thus provides an axiomatization of the similarity order in terms of conditions on the faithfulness constraints. The proposed axiomatization subsumes Tesar's concrete similarity order provided by Definition 4 as a special case.[21]

5.4.1 An initial attempt

Intuitively, a candidate has more internal similarity than another candidate provided the former incurs less faithfulness violations than the latter. This intuition is formalized by Definition 5. For the sake of generality (see below Subsection 5.5.4), the similarity order is parameterized by a set $\mathcal{F}$ of faithfulness constraints which can in principle be different from or smaller than the entire faithfulness constraint set used to define the OT typology.

Definition 5 For any candidates $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ and $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ sharing the surface form $\mathbf{d}$, let $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq^{\mathcal{F},\text{weak}}_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ provided every faithfulness constraint F in the faithfulness constraint set $\mathcal{F}$ satisfies the following inequality:

[21] Tesar defines the similarity order concretely in terms of (identity, deletion, and insertion) disparities between strings and correspondence relations in order for the resulting notion of output-drivenness to be framework-independent and thus to be able to bridge rule-based and constraint-based phonology. Yet, I submit that a *disparity* is really nothing else than a new technical term to denote a faithfulness constraint violation. Furthermore, Tesar does not shy away from correspondence relations, although they also are strictly speaking not framework-independent, but rather a representational device needed in constraint-based phonology to get around the lack of phonological derivations. Yet, Tesar (p. 34) objects that, "while in linguistics the terminology of correspondence is perhaps found most explicitly in the OT literature, the concept is equally important to any generative theory. There is a correspondence relation implicit in every SPE-style rule." I submit that the same argument applies to faithfulness constraints: even though they were only formalized in OT, faithfulness considerations are plausibly intrinsic to phonology, no matter the framework. I conclude that there is no impediment against rephrasing Tesar's definition of the similarity order in terms of faithfulness constraints. Furthermore, we will see that this restatement yields a deep understanding of the formal underpinning of the theory of output-drivenness.

(65) $F\big(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}\big) \geq F\big(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}\big)$ □

When the faithfulness constraint set $\mathcal{F}$ consists of MAX, DEP, and IDENT$_\varphi$ corresponding to features $\varphi \in \Phi$, the similarity order $\leq_{\text{sim}}^{\mathcal{F},\text{weak}}$ provided by the faithfulness-based Definition 5 coincides with the weak similarity order $\leq_{\text{sim}}^{\Phi,\text{weak}}$ provided by the concrete Definition 3 and thus inherits the drawbacks discussed in Subsection 5.3.2. A stronger axiomatization is therefore needed.

5.4.2 A stronger axiomatization

Subsection 5.2.5 has shown that the sufficient condition for output-drivenness provided by the FODC$^{\text{OT}}$ (58) is tight: any faithfulness constraint which fails the FODC$^{\text{OT}}$ yields a failure of output-drivenness. In order to amend Definition 5, I thus look at the FODC$^{\text{OT}}$. Given two candidates such that $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$, Tesar's proposition 8 requires the FODC$^{\text{OT}}$ to hold for *every* candidate $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ and *some* candidate $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$. Consider the special case where $\mathbf{c} = \mathbf{b}$ and $\rho_{\mathbf{b},\mathbf{c}}$ is the identity $\mathbb{I}_{\mathbf{b},\mathbf{b}}$ over the string $\mathbf{b} = \mathbf{c}$. Tesar's proposition thus requires in particular that there exists *some* candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ such that the first FODC$^{\text{OT}}$ implication (58a) holds, which in the specific case considered becomes (66).

(66) **If:** $F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) < F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$
Then: $F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) < F(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$

By (8), faithfulness constraints assign zero violations to the identity candidate $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$. The consequent of the implication (66) thus says that $F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) < 0$. The latter contradicts the assumption that constraint violations are non-negative. Since the consequent of the implication (66) is false, the antecedent must be false as well. This means that the similarity order $\leq_{\text{sim}}$ must be defined in such a way that $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ entails the existence of some candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ which satisfies the inequality (67).

(67) $F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \geq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$

In conclusion, the similarity order $\leq_{\text{sim}}$ must be defined in such a way that the inequality $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ entails both the intuitive condition (65) and the technical condition (67). The simplest way to achieve that is by adding the right-hand side of those two conditions yielding (68). This revised condition is stronger than the original condition (65) because of the additional term on the right-hand side of the inequality. Intuitively, this new condition (68) requires the less similar candidate $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ to "make up" not only for any faithfulness violation incurred by the more similar candidate $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ but also for any faithfulness violation incurred by the candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ which puts the two strings $\mathbf{a}$ and $\mathbf{b}$ in correspondence.

Definition 6 (Axiomatization of similarity orders) For any candidates $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ and $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ which share a surface form $\mathbf{d}$, let $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}}^{\mathcal{F}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ provided the candidate set contains a candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ which puts in correspondence the two strings a and b in such a way that every faithful constraint F in the faithfulness constraint set $\mathcal{F}$ satisfies the inequality

$$(68) \quad F(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}}) \geq F(\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}}) + \underbrace{F(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}})}_{\text{additional term}} \qquad \square$$

Magri (2016a) shows that the relation $\leq^{\mathcal{F}}_{\text{sim}}$ provided by Definition 6 is indeed a partial order on the candidate set and that it satisfies the intuitive condition (52) that identity candidates have maximal internal similarity.

To illustrate, consider the two candidates $(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}})$ and $(\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}})$ in (69a). They satisfy the similarity inequality $(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}}) \leq^{\mathcal{F}}_{\text{sim}} (\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}})$ relative to the faithfulness constraint set $\mathcal{F} = \{\textsc{Max}, \textsc{Ident}_{[\text{high}]}, \textsc{Ident}_{[\text{voice}]}\}$ because each of those three faithfulness constraints satisfies condition (68) when the candidate $(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}})$ which features in the additional term is (69b).

$$(69) \quad \text{a.} \quad \begin{array}{rl} \mathsf{a} = & \mathsf{r\ i\ d} \\ & \mathsf{|\ \ |} \\ \mathsf{d} = & \mathsf{r\ e} \end{array} \qquad \begin{array}{rl} \mathsf{b} = & \mathsf{r\ e\ d} \\ & \mathsf{|\ \ |} \\ \mathsf{d} = & \mathsf{r\ e} \end{array} \qquad \text{b.} \quad \begin{array}{rl} \mathsf{a} = & \mathsf{r\ i\ d} \\ & \mathsf{|\ \ |\ \ |} \\ \mathsf{b} = & \mathsf{r\ e\ d} \end{array}$$

For comparison, consider again the two candidates $(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}})$ and $(\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}})$ in (62a) repeated in (70a), which were found to be problematic for output-drivenness. These two candidates fail the similarity inequality $(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}}) \leq^{\mathcal{F}}_{\text{sim}} (\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}})$. In fact, the two obvious choices for the candidate $(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}})$ which features in the additional term of the revised condition (68) are (70b) and (70b′). Yet, condition (68) fails for $F = \textsc{Ident}_{[\text{voice}]}$ in the case of (70b) and for $F = \textsc{Max}$ in the case of (70b′).

$$(70) \quad \text{a.} \quad \begin{array}{rl} \mathsf{a} = & \mathsf{r\ i\ d} \\ & \mathsf{|\ \ |} \\ \mathsf{d} = & \mathsf{r\ e} \end{array} \qquad \begin{array}{rl} \mathsf{b} = & \mathsf{r\ e\ t} \\ & \mathsf{|\ \ |} \\ \mathsf{d} = & \mathsf{r\ e} \end{array} \qquad \text{b.} \quad \begin{array}{rl} \mathsf{a} = & \mathsf{r\ i\ d} \\ & \mathsf{|\ \ |\ \ |} \\ \mathsf{b} = & \mathsf{r\ e\ t} \end{array} \qquad \text{b}'. \quad \begin{array}{rl} \mathsf{a} = & \mathsf{r\ i\ d} \\ & \mathsf{|\ \ |} \\ \mathsf{b} = & \mathsf{r\ e\ t} \end{array}$$

The crucial difference between the two candidates $(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}})$ and $(\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}})$ in (69a) which satisfy the similarity inequality and those in (70a) which instead fail is that the deleted codas in the latter two candidates differ in voicing. The additional term in the revised condition (68) thus plays the role of the additional underlined condition in the deletion clause of Tesar's definition 4.

As another example, consider the two candidates $(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}})$ and $(\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}})$ in (71a). They satisfy the similarity inequality $(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}}) \leq^{\mathcal{F}}_{\text{sim}} (\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}})$ relative to the faithfulness constraint set $\mathcal{F} = \{\textsc{Max}, \textsc{Ident}_{[\text{high}]}, \textsc{Ident}_{[\text{place}]}\}$ because each of those faithfulness constraints satisfies condition (68) when the candidate $(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}})$ which features in the additional term is (71b).

$$(71) \quad \text{a.} \quad \begin{array}{rl} \mathsf{a} = & \mathsf{p\ a} \\ & \mathsf{|\ \ |} \\ \mathsf{d} = & \mathsf{t\ i} \end{array} \qquad \begin{array}{rl} \mathsf{b} = & \mathsf{p\ i} \\ & \mathsf{|\ \ |} \\ \mathsf{d} = & \mathsf{t\ i} \end{array} \qquad \text{b.} \quad \begin{array}{rl} \mathsf{a} = & \mathsf{p\ a} \\ & \mathsf{|\ \ |} \\ \mathsf{b} = & \mathsf{p\ i} \end{array}$$

For comparison, consider again the two candidates $(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}})$ and $(\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}})$ in (62b) repeated in (72a), which were found to be problematic for output-drivenness. Those two candidates fail the similarity inequality $(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}}) \leq^{\mathcal{F}}_{\text{sim}} (\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}})$. In fact, the two obvious choices for the candidate $(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}})$ which features in the additional term of the revised condition (68) are (72a) and (72b). Yet, condition (68) fails for $F = \text{IDENT}_{[\text{place}]}$ in the case of (72a) and it fails for $F = \text{MAX}$ in the case of (72b′).

(72) a. $\begin{array}{rl} \mathbf{a} = & \mathsf{p\,a} \\ & \mid\;\mid \\ \mathbf{d} = & \mathsf{t\,i} \end{array}$ $\quad\begin{array}{rl} \mathbf{b} = & \mathsf{k\,i} \\ & \mid\;\mid \\ \mathbf{d} = & \mathsf{t\,i} \end{array}$ b. $\begin{array}{rl} \mathbf{a} = & \mathsf{p\,a} \\ & \mid\;\mid \\ \mathbf{b} = & \mathsf{k\,i} \end{array}$ b′. $\begin{array}{rl} \mathbf{a} = & \mathsf{p\,a} \\ & \;\;\;\mid \\ \mathbf{b} = & \mathsf{k\,i} \end{array}$

The crucial difference between the two candidates $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ and $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ in (71a) which satisfy the similarity inequality and those in (72a) which instead fail is that the underlying onsets in the latter two candidates differ in place not only with their surface correspondent but also between each other. The additional term in the revised condition (68) thus plays the role of the additional underlined condition in the identity clause of Tesar's definition 4.[22]

5.4.3 The axiomatic definition subsumes Tesar's concrete definition

Assume that the faithfulness constraint set $\mathcal{F}$ consists of the faithfulness constraints that Tesar focuses on, namely MAX, DEP, and IDENT_{φ} for any feature φ in the feature set Φ. The sophisticated analysis developed in Tesar (2013, chapter 3) can be rebooted (see Magri 2016a for details) to establish that:

Proposition 9 *For any two candidates $(\boldsymbol{a}, \boldsymbol{d}, \rho_{\boldsymbol{a},\boldsymbol{d}})$ and $(\boldsymbol{b}, \boldsymbol{d}, \rho_{\boldsymbol{b},\boldsymbol{d}})$, if the similarity inequality $(\boldsymbol{a}, \boldsymbol{d}, \rho_{\boldsymbol{a},\boldsymbol{d}}) \leq^{\Phi}_{sim} (\boldsymbol{b}, \boldsymbol{d}, \rho_{\boldsymbol{b},\boldsymbol{d}})$ holds relative the similarity order $\leq^{\Phi}_{sim}$ provided by Tesar's concrete Definition 4, then the similarity inequality $(\boldsymbol{a}, \boldsymbol{d}, \rho_{\boldsymbol{a},\boldsymbol{d}}) \leq^{\mathcal{F}}_{sim} (\boldsymbol{b}, \boldsymbol{d}, \rho_{\boldsymbol{b},\boldsymbol{d}})$ holds relative to the similarity order $\leq^{\mathcal{F}}_{sim}$ provided by the axiomatic Definition 6.* □

Yet, the two similarity orders $\leq^{\Phi}_{\text{sim}}$ and $\leq^{\mathcal{F}}_{\text{sim}}$ are not equivalent: it is easy to construct cases where $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq^{\mathcal{F}}_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ but $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \not\leq^{\Phi}_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$, even when $\mathcal{F}$ only consists of the three types of faithfulness constraints that Tesar focuses on (see Magri 2016a for discussion). In conclusion, the proposed

[22] Consider again the candidates $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ and $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ in (55), repeated below in (ia) and (ib). In order to secure the argument made in Subsection 5.1.3, we need to secure the similarity inequality $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq^{\mathcal{F}}_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ with $\mathcal{F} = \{\text{IDENT}_{[\text{voice}]}, \text{IDENT}_{[\text{cont}]}\}$. Indeed, condition (68) holds for both faithfulness constraints in $\mathcal{F}$ when the candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ in the additional term of (68) is defined as in (ic).

(i) a. $\begin{array}{rl} \mathbf{a} = & \mathsf{p} \\ & \mid \\ \mathbf{d} = & \beta \end{array}$ b. $\begin{array}{rl} \mathbf{b} = & \mathsf{b} \\ & \mid \\ \mathbf{d} = & \beta \end{array}$ c. $\begin{array}{rl} \mathbf{a} = & \mathsf{p} \\ & \mid \\ \mathbf{b} = & \mathsf{b} \end{array}$

Consider next the candidates $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ and $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ in (56), repeated below in (iia) and (iib). In order to secure the argument made in Subsection 5.2.1, we need to secure the similarity inequality $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq^{\mathcal{F}}_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ with $\mathcal{F} = \{\text{IDENT}_{[\text{voice}]}, \text{IDENT}_{[\text{low}]}\}$. Indeed, condition (68) holds for both faithfulness constraints in $\mathcal{F}$ when the candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ in the additional term in (68) is defined as in (iic).

(ii) a. $\begin{array}{rl} \mathbf{a} = & \mathsf{t\,e\,t\,d} \\ & \mid\;\mid\;\mid\!/ \\ \mathbf{d} = & \mathsf{t\,a\,t} \end{array}$ b. $\begin{array}{rl} \mathbf{b} = & \mathsf{t\,a\,d} \\ & \mid\;\mid\;\mid \\ \mathbf{d} = & \mathsf{t\,a\,t} \end{array}$ c. $\begin{array}{rl} \mathbf{a} = & \mathsf{t\,e\,t\,d} \\ & \mid\;\mid\;\;/ \\ \mathbf{b} = & \mathsf{t\,a\,d} \end{array}$

axiomatization of the similarity order subsumes Tesar's concrete definition as a special, concrete case. The notion of output-drivenness relative to the axiomatized similarity order $\leq^{\mathcal{F}}_{\text{sim}}$ is thus slightly stronger than Tesar's original notion (because it holds relative to a similarity order which is slightly looser) and the resulting theory therefore slightly more general.

5.4.4 The composition candidate and the FODC$^{OT}_{comp}$

Consider two candidates $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ and $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ which satisfy the similarity inequality $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq^{\mathcal{F}}_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$. Tesar's Proposition 8 considers an arbitrary third candidate $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ which puts the more similar underlying form $\mathbf{b}$ in correspondence with a candidate $\mathbf{c}$ different from $\mathbf{d}$ through some relation $\rho_{\mathbf{b},\mathbf{c}}$. The proposition then requires the existence of a fourth candidate $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$ which puts the less similar underlying form $\mathbf{a}$ in correspondence with that same candidate $\mathbf{c}$ through some relation $\rho_{\mathbf{a},\mathbf{c}}$, as represented in (73).

(73)
$$
\begin{tikzcd}
\mathbf{a} \arrow[r, "\rho_{\mathbf{a},\mathbf{d}}"] \arrow[dr, "\rho_{\mathbf{a},\mathbf{b}}"'] \arrow[drr, bend right=60, "\rho_{\mathbf{a},\mathbf{c}}"] & \mathbf{d} & \\
 & \mathbf{b} \arrow[u, "\rho_{\mathbf{b},\mathbf{d}}"'] \arrow[r, "\rho_{\mathbf{b},\mathbf{c}}"] & \mathbf{c}
\end{tikzcd}
$$

By Definition 6 of the similarity order $\leq^{\mathcal{F}}_{\text{sim}}$, the assumption $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq^{\mathcal{F}}_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ ensures the existence of a correspondence relation $\rho_{\mathbf{a},\mathbf{b}}$ between the two strings $\mathbf{a}$ and $\mathbf{b}$. It is thus natural to assume that $\rho_{\mathbf{a},\mathbf{c}}$ is the composition $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$ of $\rho_{\mathbf{a},\mathbf{b}}$ and $\rho_{\mathbf{b},\mathbf{c}}$. The existence of this composition candidate is guaranteed by the transitivity axiom (6). The FODC$^{\text{OT}}$ (58) can thus be specialized as in (74), which will be referred to as the FODC$^{\text{OT}}_{\text{comp}}$.

(74) a. **If:** $F\big(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}\big) < F(\mathbf{a}, \mathbf{c}, \overbrace{\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}}^{\rho_{\mathbf{a},\mathbf{c}}})$
Then: $F\big(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}\big) < F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$

b. **If:** $F\big(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}\big) < F\big(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}\big)$
Then: $F\big(\mathbf{a}, \mathbf{c}, \underbrace{\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}}_{\rho_{\mathbf{a},\mathbf{c}}}\big) < F\big(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}\big)$

5.5 OT output-drivenness and the triangle inequality

The condition for output-drivenness provided by the FODC$^{\text{OT}}_{\text{comp}}$ (74) looks like a technical condition without an intuitive interpretation. The rest of this section will derive an intuitive metric interpretation of this condition by investigating its relationship with the triangle inequality.

5.5.1 The FTI_{comp} entails the $FODC^{OT}_{comp}$ in the general case

Consider two candidates $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ and $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ which satisfy the similarity inequality $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq^{\mathcal{F}}_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ relative to the similarity order provided by Definition 6. This means that there exists a candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ such that each faithfulness constraint F in the faithfulness constraint set $\mathcal{F}$ satisfies the inequality (68), repeated in (75).

(75) $F\big(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}\big) \geq F\big(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}\big) + F\big(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}\big)$

Assume now that the faithfulness constraint F measures the phonological distance between underlying and surface forms in compliance with the metric triangle inequality, which has been formalized in Section 2 as the FTI_{comp}. The latter boils down to the inequality (76) in the specific case of the two candidates $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ and $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ and their composition candidate $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$.

(76) $F\big(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}\big) \leq F\big(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}\big) + F\big(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}\big)$

The inequality (75) allows $F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ in (76) to be upper bounded with $F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) - F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$, yielding the inequality (77).

(77) $F\big(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}\big) - F\big(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}\big) \leq F\big(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}\big) - F\big(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}\big)$

Assume that the antecedent of the first $\text{FODC}^{\text{OT}}_{\text{comp}}$ implication (74a) holds, namely that $F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) < F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$. The inequality (77) then entails that $F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) < F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$, thus ensuring that the consequent of the implication holds as well. In other words, the inequality (77) entails the first $\text{FODC}^{\text{OT}}_{\text{comp}}$ implication (74a). An analogous reasoning applies to the second $\text{FODC}^{\text{OT}}_{\text{comp}}$ implication (74b). In conclusion, the FTI_{comp} entails Tesar's $\text{FODC}^{\text{OT}}_{\text{comp}}$ relative to the similarity order provided by Definition 6, as summarized in the following Proposition 10, which is the second main technical result of this paper. No assumptions are made on the nature of the faithfulness constraints (for instance, they are not required to be categorical) or on the correspondence relations in the candidate set (for instance, they are not required to be one-to-one).

Proposition 10 *If a faithfulness constraint F satisfies the FTI_{comp} and it belongs to the faithfulness constraint set $\mathcal{F}$ used define the similarity order $\leq^{\mathcal{F}}_{sim}$ according to Definition 6, then F satisfies the $FODC^{OT}_{comp}$ (74) relative to that similarity order $\leq^{\mathcal{F}}_{sim}$.* □

5.5.2 The FTI_{comp} is stronger than the $FODC^{OT}_{comp}$ in the general case

Proposition 10 says that the FTI_{comp} entails the $\text{FODC}^{\text{OT}}_{\text{comp}}$. The following counterexample shows that the reverse entailment fails in the general case, so that the FTI_{comp} is stronger than the $\text{FODC}^{\text{OT}}_{\text{comp}}$. Let $\ell(\mathbf{a})$ be the length of the string $\mathbf{a}$. Assume that the candidate set displays no epenthesis and thus consists of candidates $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ whose surface form $\mathbf{b}$ is not longer than the the underlying form $\mathbf{a}$: $\ell(\mathbf{b}) \leq \ell(\mathbf{a})$. Since correspondence relations play no role in the counterexample, I omit them and represent candidates simply as pairs of strings. Consider the faithfulness constraint (78) which assigns to a candidate $(\mathbf{a}, \mathbf{b})$ a number of violations equal to the squared difference of the length of its two strings. This is a proper faithfulness constraint, in the sense that it assigns zero violations to any identity candidate, in compliance with the definitional faithfulness condition (8).

(78) $F(\mathbf{a}, \mathbf{b}) = \big(\ell(\mathbf{a}) - \ell(\mathbf{b})\big)^2$

This faithfulness constraint F satisfies both $\text{FODC}^{\text{OT}}_{\text{comp}}$ implications (74), as their antecedent and consequent are equivalent.[23] Yet, F fails at the FTI_{comp} for any two candidates $(\mathbf{a}, \mathbf{b})$ and $(\mathbf{b}, \mathbf{c})$ and their composition candidate $(\mathbf{a}, \mathbf{c})$.[24] This constraint F thus shows that the $\text{FODC}^{\text{OT}}_{\text{comp}}$ is weaker than the FTI_{comp}.

5.5.3 The FTI_{comp}, $FODC^{OT}_{comp}$, and FIC^{OT}_{comp} are equivalent for categorical constraints

Subsections 5.2.3 and 5.5.1 have shown that the FTI_{comp} entails the $\text{FODC}^{\text{OT}}_{\text{comp}}$ which in turn entails the $\text{FIC}^{\text{OT}}_{\text{comp}}$. Assume now that all correspondence relations in the candidate set are one-to-one, as otherwise output-drivenness fails (as shown in Subsection 5.2.1) and the $\text{FODC}^{\text{OT}}_{\text{comp}}$ must therefore fail as well. Assume furthermore that each faithfulness constraint in $\mathcal{F}$ satisfies McCarthy's (strengthened) categoricity conjecture, formalized in Subsection 2.3. Under these assumptions, the $\text{FIC}^{\text{OT}}_{\text{comp}}$ entails the FTI_{comp}, as shown in Subsection 3.3.3. I conclude that the FTI_{comp}, the $\text{FODC}^{\text{OT}}_{\text{comp}}$, and the $\text{FIC}^{\text{OT}}_{\text{comp}}$ are equivalent, as summarized in the following proposition.

Proposition 11 *Assume that the correspondence relations in the candidate set are all one-to-one. Consider a faithfulness constraint F which is C-categorical; or I-categorical and O-monotone; or O-categorical and I-monotone. Assume that F belongs to the faithfulness subset $\mathcal{F}$ used to define the similarity order $\leq^{\mathcal{F}}_{sim}$ according to Definition 6. The constraint F satisfies the FTI_{comp} if and only if it satisfies the $FODC^{OT}_{comp}$ if and only if it satisfies the FIC^{OT}_{comp}.* □

Proposition 11 provides an intuitive interpretation of the rather technical sufficient condition for OT output-drivenness provided by the $\text{FODC}^{\text{OT}}_{\text{comp}}$. In

[23] Consider the first $\text{FODC}^{\text{OT}}_{\text{comp}}$ implication (74a). Its antecedent and consequent are shown to be equivalent in (i), using the fact that $x < y$ iff $x^2 < y^2$, for any $x, y \geq 0$

(i)
$$\begin{aligned} F(\mathbf{a}, \mathbf{d}) < F(\mathbf{a}, \mathbf{c}) &\iff (\ell(\mathbf{a}) - \ell(\mathbf{d}))^2 < (\ell(\mathbf{a}) - \ell(\mathbf{c}))^2 \\ &\iff (\ell(\mathbf{a}) - \ell(\mathbf{d})) < (\ell(\mathbf{a}) - \ell(\mathbf{c})) \\ &\iff -\ell(\mathbf{d}) < -\ell(\mathbf{c}) \\ &\iff (\ell(\mathbf{b}) - \ell(\mathbf{d})) < (\ell(\mathbf{b}) - \ell(\mathbf{c})) \\ &\iff (\ell(\mathbf{b}) - \ell(\mathbf{d}))^2 < (\ell(\mathbf{b}) - \ell(\mathbf{c}))^2 \\ &\iff F(\mathbf{b}, \mathbf{d}) < F(\mathbf{b}, \mathbf{c}) \end{aligned}$$

An analogous reasoning holds for the second $\text{FODC}^{\text{OT}}_{\text{comp}}$ implication (74b).

[24] In fact $F(\mathbf{a}, \mathbf{c}) = (\ell(\mathbf{a}) - \ell(\mathbf{c}))^2 = [(\ell(\mathbf{a}) - \ell(\mathbf{b})) + (\ell(\mathbf{b}) - \ell(\mathbf{c}))]^2 \geq (\ell(\mathbf{a}) - \ell(\mathbf{b}))^2 + (\ell(\mathbf{b}) - \ell(\mathbf{c}))^2 = F(\mathbf{a}, \mathbf{b}) + F(\mathbf{b}, \mathbf{c})$, contradicting the FTI_{comp}.

fact, it says that the $\mathrm{FODC}^{\mathrm{OT}}_{\mathrm{comp}}$ simply requires (categorical and monotone) faithfulness constraints to measure the phonological distance between underlying and surface forms in compliance with the metrical axiom of the triangle inequality, as formalized through the $\mathrm{FTI}_{\mathrm{comp}}$. Furthermore, Proposition 11 provides a straightforward characterization of the faithfulness constraints which satisfy the $\mathrm{FODC}^{\mathrm{OT}}_{\mathrm{comp}}$. In fact, the faithfulness constraints listed by Proposition 2 in Section 3 as satisfying the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ are all categorical and monotone. The equivalence between the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ and the $\mathrm{FODC}^{\mathrm{OT}}_{\mathrm{comp}}$ established by Proposition 11 thus ensures that they also all satisfy the $\mathrm{FODC}^{\mathrm{OT}}_{\mathrm{comp}}$ (under the assumption that all correspondence relations are one-to-one).

5.5.4 The restriction to faithfulness constraints which belong to the subset $\mathcal{F}$

Consider again the two candidates $(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}}) = (/\mathsf{p}/, [\beta])$ and $(\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}}) = (/\mathsf{b}/, [\beta])$ in (55) (correspondence relations play no role because I am considering singleton segments). As discussed in Subsection 5.1.3, any plausible measure of internal similarity should yield that candidate $(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}})$ has less internal similarity than candidate $(\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}})$, because the former candidate displays disparities for both voicing and continuancy while the latter displays a disparity for continuancy but not for voicing. Consider the faithfulness constraint $F = \textsc{Ident}_{[\text{voice}]} \vee \textsc{Ident}_{[\text{cont}]}$ which is the disjunction of the two identity faithfulness constraints $\textsc{Ident}_{[\text{voice}]}$ and $\textsc{Ident}_{[\text{cont}]}$ for voicing and continuancy. Subsection 5.2.4 has shown that F fails at the $\mathrm{FODC}^{\mathrm{OT}}_{\mathrm{comp}}$. Yet, Proposition 2b in Subsection 3.2.4 states that F succeeds at the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$. Furthermore, F is obviously C-categorical, as it is the disjunction of two identity constraints which are both C-categorical. The fact that F satisfies the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ but not the $\mathrm{FODC}^{\mathrm{OT}}_{\mathrm{comp}}$ is *not* a counterexample to the entailment from the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ to the $\mathrm{FODC}^{\mathrm{OT}}_{\mathrm{comp}}$ ensured by Proposition 11. In fact, that proposition only looks at the faithfulness constraints which crucially belong to the faithfulness constraint set $\mathcal{F}$ used to define the similarity order $\leq^{\mathcal{F}}_{\mathrm{sim}}$ according to Definition 6. Crucially, the disjunctive faithfulness constraint $F = \textsc{Ident}_{[\text{voice}]} \vee \textsc{Ident}_{[\text{cont}]}$ cannot belong to the set $\mathcal{F}$. In fact, the inequality (79) required by Definition 6 for the similarity inequality $(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}}) \leq^{\mathcal{F}}_{\mathrm{sim}} (\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}})$ to hold fails for this specific faithfulness constraint F.[25] In other words, if F did belong to $\mathcal{F}$, the two candidates $(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}})$ and $(\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}})$ would not be comparable relative to the similarity order $\leq^{\mathcal{F}}_{\mathrm{sim}}$, contrary to intuitions.

(79) $$\underbrace{F(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}})}_{=F(/\mathsf{p}/,\,[\beta])=1} \not\geq \underbrace{F(\mathsf{b}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{b}})}_{=F(/\mathsf{b}/,\,[\beta])=1} + \underbrace{F(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}})}_{=F(/\mathsf{p}/,\,[\mathsf{b}])=1}$$

The reason why Definition 6 of the similarity order $\leq^{\mathcal{F}}_{\mathrm{sim}}$ has been parameterized by a faithfulness constraint subset $\mathcal{F}$ possibly smaller than the entire

[25] I am assuming that the correspondence relation $\rho_{\mathsf{a},\mathsf{b}}$ in the candidate $(\mathsf{a}, \mathsf{b}, \rho_{\mathsf{a},\mathsf{b}}) = (/\mathsf{p}/, [\mathsf{b}])$ does put the singleton underlying and surface segments in correspondence. If that is not the case, then the inequality (79) would indeed succeed for $F = \textsc{Ident}_{[\text{voice}]} \vee \textsc{Ident}_{[\text{cont}]}$ but it would fail for $F = \textsc{Max}$ and $F = \textsc{Dep}$.

faithfulness constraint set is precisely to allow for the possibility that certain faithfulness constraints (and in particular those derived from other more basic faithfulness constraints through operations such as local disjunction) not be considered in the computation of similarity.

5.6 Summary

This section has reconstructed Tesar's (2013) theory of output-drivenness. The sufficient condition for output-drivenness provided by his FODC$^{\mathrm{OT}}$ has been shown to admit a metric interpretation: it effectively requires the faithfulness constraints (which belong to the faithfulness constraint set $\mathcal{F}$ used to compare similarity) to measure the phonological distance between underlying and surface forms in compliance with the metric triangle inequality. This interpretation holds under McCarthy's (strengthened) categoricity conjecture. Because of this reinterpretation, the FODC$^{\mathrm{OT}}$ turns out to be equivalent to the sufficient condition for idempotency provided by the FIC$^{\mathrm{OT}}$ (for faithfulness constraints which are categorical and belong to $\mathcal{F}$).

6 Output-drivenness in Harmonic Grammar

This section extends the theory of output-drivenness from OT to HG. This extension will provide a more pristine view of the relationship between output-drivenness and the triangle inequality, which does not rely on categoricity.

6.1 A condition on the faithfulness constraint set: the FODC$^{\mathrm{HG}}_{\mathrm{comp}}$

Subsection 5.2.2 has recalled Tesar's (2013) Proposition 8, which provides guarantees for the output-drivenness of the grammars in an OT typology under the assumption that each faithfulness constraint satisfies the FODC$^{\mathrm{OT}}$ (58). The following Proposition 12 provides the HG analogue of Tesar's OT Proposition 8. The proof is analogous to that of Proposition 5 for HG idempotency and it is provided in Appendix A.3. The implication (80) will be referred to as the *HG faithfulness output-drivenness condition* (FODC$^{\mathrm{HG}}$). Proposition 12 makes no assumptions on the markedness constraints, on the nature of the faithfulness constraints (for instance, it does not require them to be categorical), on the correspondence relations in the candidate set (for instance, it does not require them to be one-to-one), or on the similarity order $\leq_{\mathrm{sim}}$ (which is indeed left completely unspecified).

Proposition 12 *Assume that, for any two candidates* $(\boldsymbol{a}, \boldsymbol{d}, \rho_{\boldsymbol{a},\boldsymbol{d}})$ *and* $(\boldsymbol{b}, \boldsymbol{d}, \rho_{\boldsymbol{b},\boldsymbol{d}})$ *such that* $(\boldsymbol{a}, \boldsymbol{d}, \rho_{\boldsymbol{a},\boldsymbol{d}}) \leq_{sim} (\boldsymbol{b}, \boldsymbol{d}, \rho_{\boldsymbol{b},\boldsymbol{d}})$*, for every other candidate* $(\boldsymbol{b}, \boldsymbol{c}, \rho_{\boldsymbol{b},\boldsymbol{c}})$ *different from* $(\boldsymbol{b}, \boldsymbol{d}, \rho_{\boldsymbol{b},\boldsymbol{d}})$*, the candidate set also contains a candidate* $(\boldsymbol{a}, \boldsymbol{c}, \rho_{\boldsymbol{a},\boldsymbol{c}})$ *different from* $(\boldsymbol{a}, \boldsymbol{d}, \rho_{\boldsymbol{a},\boldsymbol{d}})$ *such that the following condition (80) holds for any faithfulness constraint* F *in the constraint set.*

(80) *For every choice of the constant ξ (with no restrictions on its sign):*
If: $F\big(\boldsymbol{b}, \boldsymbol{c}, \rho_{\boldsymbol{b},\boldsymbol{c}}\big) \leq F\big(\boldsymbol{b}, \boldsymbol{d}, \rho_{\boldsymbol{b},\boldsymbol{d}}\big) + \xi$
Then: $F\big(\boldsymbol{a}, \boldsymbol{c}, \rho_{\boldsymbol{a},\boldsymbol{c}}\big) \leq F\big(\boldsymbol{a}, \boldsymbol{d}, \rho_{\boldsymbol{a},\boldsymbol{d}}\big) + \xi$

Then, the HG grammar corresponding to any weighting of the constraint set is output-driven relative to the similarity order $\leq_{sim}$. □

Let's now consider the special case where the similarity order is defined as in Subsection 5.4.2. The similarity condition $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq^{\mathcal{F}}_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ thus means that there exists some candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ which puts the two underlying forms $\mathbf{a}$ and $\mathbf{b}$ in correspondence and validates the crucial inequality (68) for every faithfulness constraint F in the faithfulness constraint subset $\mathcal{F}$. By reasoning as in Subsection 5.4.4, the FODC$^{\text{HG}}$ (80) can be specialized to the FODC$^{\text{HG}}_{\text{comp}}$ (81), by assuming that the candidate $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$ which appears on the left-hand side of the inequality in the consequent is the composition candidate $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$ of the two candidates $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ and $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$, whose existence is guaranteed by the transitivity axiom (6).

(81) For every choice of the constant ξ (with no restrictions on its sign):
If: $F\big(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}\big) \leq F\big(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}\big) + \xi$
Then: $F\big(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}\big) \leq F\big(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}\big) + \xi$

The sufficient condition for output-drivenness provided by the FODC$^{\text{HG}}_{\text{comp}}$ is *tight*: any faithfulness constraint which fails the FODC$^{\text{HG}}_{\text{comp}}$ admits an elementary counterexample where output-drivenness fails.

6.2 HG output-drivenness and the triangular inequality

This section shows that the FODC$^{\text{HG}}_{\text{comp}}$ is stronger than the FODC$^{\text{OT}}_{\text{comp}}$ so that the FODC$^{\text{HG}}_{\text{comp}}$ is equivalent to the triangular inequality for any faithfulness constraint, while we have seen that the equivalence holds for the FODC$^{\text{OT}}_{\text{comp}}$ only under McCarthy's (strengthened) categoricity conjecture.

6.2.1 The FODC^{HG} entails the FIC^{HG} in the general case

Subsection 5.2.3 has shown that the sufficient condition for OT output-drivenness provided by the FODC$^{\text{OT}}$ entails the sufficient condition for OT idempotency provided by the FIC$^{\text{OT}}$, because the latter coincides with the former in the special case where $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ is the identity candidate. An analogous reasoning holds in the case of HG. The sufficient condition for HG output-drivenness provided by the FODC$^{\text{HG}}$ (80) reduces to (82) when $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ is the identity candidate. The latter is the sufficient condition for HG idempotency provided by the FIC$^{\text{HG}}$ (45), as derived in Section 4.1.3.[26]

[26] As explained in footnote 17, it makes no difference whether the constant ξ in (45)/(82) is restricted to be non-negative or allowed to be negative. That is of course not the case for (80): the fact that ξ is allowed to be negative makes it a stronger condition.

(82) For every choice of the constant ξ (with no restrictions on its sign):
If: $F\big(\mathsf{b}, \mathsf{c}, \rho_{\mathsf{b},\mathsf{c}}\big) \leq \xi$
Then: $F\big(\mathsf{a}, \mathsf{c}, \rho_{\mathsf{a},\mathsf{c}}\big) \leq F\big(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}}\big) + \xi$

In conclusion, the FODC$^{\mathrm{HG}}$ entails the FIC$^{\mathrm{HG}}$ in the general case, no matter how the similarity order in the definition of output-drivenness is defined. This conclusion matches the fact that output-drivenness entails idempotency independently of the grammatical fromework, as noted in Subsection 5.1.2.

6.2.2 The FODC^{HG} entails the FODC^{OT} in the general case

Subsection 5.2.2 has introduced Tesar's sufficient condition FODC$^{\mathrm{OT}}$ for OT output-drivenness, repeated in (83) for ease of comparison.

(83) a. **If:** $F\big(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}}\big) < F\big(\mathsf{a}, \mathsf{c}, \rho_{\mathsf{a},\mathsf{c}}\big)$
Then: $F\big(\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}}\big) < F\big(\mathsf{b}, \mathsf{c}, \rho_{\mathsf{b},\mathsf{c}}\big)$

b. **If:** $F\big(\mathsf{b}, \mathsf{c}, \rho_{\mathsf{b},\mathsf{c}}\big) < F\big(\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}}\big)$
Then: $F\big(\mathsf{a}, \mathsf{c}, \rho_{\mathsf{a},\mathsf{c}}\big) < F\big(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}}\big)$

These two implications (83a) and (83b) are closely related: if we were to replace "$<$" with "$\leq$", then one would be the counter-positive of the other and we could retain only one of the two implications. This intuition is brought out by the following Proposition 13: the two implications (83a) and (83b) are jointly equivalent to the condition (84). The latter indeed coincides with the second implication (83b) where "$<$" has been replaced with "$\leq$" at the price of adding a small constant ξ at the right-hand side. The proof of this equivalence is straightforward but tedious, and it is therefore relegated to Appendix A.4.

Proposition 13 *The two FODC^{OT} implications (83) are jointly equivalent to the following condition:*

(84) For every choice of the constant ξ between -1 and $+1$ (both excluded):
If: $F\big(\boldsymbol{b}, \boldsymbol{c}, \rho_{\boldsymbol{b},\boldsymbol{c}}\big) \leq F\big(\boldsymbol{b}, \mathrm{d}, \rho_{\boldsymbol{b},\boldsymbol{d}}\big) + \xi$
Then: $F\big(\boldsymbol{a}, \boldsymbol{c}, \rho_{\boldsymbol{a},\boldsymbol{c}}\big) \leq F\big(\boldsymbol{a}, \boldsymbol{d}, \rho_{\boldsymbol{a},\boldsymbol{d}}\big) + \xi$ □

The restatement (84) of the FODC$^{\mathrm{OT}}$ shows that it is actually entailed by the sufficient condition FODC$^{\mathrm{HG}}$ (80) for HG output-drivenness in the general case, matching the fact that HG typologies can be larger than OT typologies.

6.2.3 The FODC$^{HG}_{comp}$, FIC^{HG}, and FTI$_{comp}$ are equivalent in the general case

The FODC$^{\mathrm{HG}}_{\mathrm{comp}}$ entails the FIC$^{\mathrm{HG}}_{\mathrm{comp}}$ in the general case (as seen in Subsection 6.2.1). The FIC$^{\mathrm{HG}}_{\mathrm{comp}}$ is in turn equivalent to the triangle inequality formalized though the FTI$_{\mathrm{comp}}$ (as seen in Subsection 4.2.2). Thus, the FODC$^{\mathrm{HG}}_{\mathrm{comp}}$ entails the FTI$_{\mathrm{comp}}$. This entailment holds under no assumptions on the faithfulness constraints. To investigate the reverse entailment, consider a faithfulness constraint F which belongs to the faithfulness constraint set $\mathcal{F}$ used to define the similarity order $\leq^{\mathcal{F}}_{\mathrm{sim}}$ according to Definition 6 in Subsection 5.4.2. Consider two candidates such that $(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}}) \leq^{\mathcal{F}}_{\mathrm{sim}} (\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}})$. As seen in Subsection 5.5.1, the $\mathrm{FTI}_{\mathrm{comp}}$ entails the inequality (77), repeated in (85). And the latter in turn straightforwardly entails the $\mathrm{FODC}^{\mathrm{HG}}_{\mathrm{comp}}$.

(85) $F\big(\mathsf{a}, \mathsf{c}, \rho_{\mathsf{a},\mathsf{b}}\rho_{\mathsf{b},\mathsf{c}}\big) - F\big(\mathsf{b}, \mathsf{c}, \rho_{\mathsf{b},\mathsf{c}}\big) \leq F\big(\mathsf{a}, \mathsf{d}, \rho_{\mathsf{a},\mathsf{d}}\big) - F\big(\mathsf{b}, \mathsf{d}, \rho_{\mathsf{b},\mathsf{d}}\big)$

In conclusion, the $\mathrm{FODC}^{\mathrm{HG}}_{\mathrm{comp}}$ relative to the similarity order $\leq^{\mathcal{F}}_{\mathrm{sim}}$ provided by the Definition 6 is equivalent to both the $\mathrm{FIC}^{\mathrm{HG}}_{\mathrm{comp}}$ and the $\mathrm{FTI}_{\mathrm{comp}}$, as summarized in the following Proposition 10. This equivalence holds for any faithfulness constraint which belongs to $\mathcal{F}$. No additional assumptions are made neither on the nature of the faithfulness constraints (for instance, they are not required to be categorical) nor on the correspondence relations in the candidate set (for instance, they are not required to be one-to-one).

Proposition 14 *The $FODC^{HG}_{comp}$ relative to the similarity order $\leq^{\mathcal{F}}_{sim}$ is equivalent to both the FIC^{HG}_{comp} and the FTI_{comp} for any faithfulness constraint which belongs to the faithfulness set $\mathcal{F}$ used to define the similarity order.* □

This proposition provides an intuitive interpretation of the rather technical sufficient condition for HG output-drivenness provided by the $\mathrm{FODC}^{\mathrm{HG}}_{\mathrm{comp}}$. In fact, it says that the $\mathrm{FODC}^{\mathrm{HG}}_{\mathrm{comp}}$ simply requires the faithfulness constraints (which belong to $\mathcal{F}$) to measure the phonological distance between underlying and surface forms in compliance with the metrical axiom of the triangle inequality, as formalized through the $\mathrm{FTI}_{\mathrm{comp}}$.

6.2.4 The $FODC^{HG}_{comp}$, $FODC^{OT}_{comp}$ and FIC^{OT}_{comp} are equivalent for categorical constraints

The $\mathrm{FODC}^{\mathrm{HG}}_{\mathrm{comp}}$ entails the $\mathrm{FODC}^{\mathrm{OT}}_{\mathrm{comp}}$ in the general case (as seen in Subsection 6.2.2). The $\mathrm{FODC}^{\mathrm{OT}}_{\mathrm{comp}}$ in turn entails the sufficient condition for OT idempotency provided by the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ (as seen in Subsection 5.2.3). Thus, the $\mathrm{FODC}^{\mathrm{HG}}_{\mathrm{comp}}$ entails the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$. This entailment holds under no assumptions on the faithfulness constraints. To investigate the reverse entailment, assume that all correspondence relations in the candidate set are one-to-one, as otherwise output-drivenness fails (as seen in Subsection 5.2.1) and the $\mathrm{FODC}^{\mathrm{HG}}_{\mathrm{comp}}$ thus fails as well. Consider a faithfulness constraint which satisfies McCarthy's (strengthened) categoricity conjecture, formalized in Subsection 2.3. Under these assumptions, the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ entails the $\mathrm{FTI}_{\mathrm{comp}}$ (as seen in Subsection 3.3.3). The $\mathrm{FTI}_{\mathrm{comp}}$ in turn entails the $\mathrm{FODC}^{\mathrm{HG}}_{\mathrm{comp}}$ for every faithfulness constraint which belongs to the faithfulness constraint set $\mathcal{F}$ used to define the similarity order $\leq^{\mathcal{F}}_{\mathrm{sim}}$ (as seen in Subsection 6.2.3). We have thus obtained the following equivalence among the $\mathrm{FODC}^{\mathrm{HG}}_{\mathrm{comp}}$, the $\mathrm{FODC}^{\mathrm{OT}}_{\mathrm{comp}}$ and the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$.

Proposition 15 *Assume that the correspondence relations in the candidate set are all one-to-one. Consider a faithfulness constraint F which is C-categorical; or I-categorical and O-monotone; or O-categorical and I-monotone. Assume* *that F belongs to the faithfulness set $\mathcal{F}$ used to define the similarity order $\leq^{\mathcal{F}}_{sim}$ according to Definition 6. The constraint F satisfies the FODC$^{HG}_{comp}$ if and only if it satisfies the FODC$^{OT}_{comp}$ if and only if it satisfies the FIC$^{OT}_{comp}$.* $\square$

This Proposition 11 provides a straightforward characterization of the faithfulness constraints which satisfy the FODC$^{\mathrm{HG}}_{\mathrm{comp}}$. In fact, the faithfulness constraints listed by Proposition 2 in Section 3 as satisfying the FIC$^{\mathrm{OT}}_{\mathrm{comp}}$ are all categorical and monotone. The equivalence between the FIC$^{\mathrm{OT}}_{\mathrm{comp}}$ and the FODC$^{\mathrm{HG}}_{\mathrm{comp}}$ established by Proposition 15 thus ensures that they also all satisfy the FODC$^{\mathrm{HG}}_{\mathrm{comp}}$ relative to $\leq^{\mathcal{F}}_{\mathrm{sim}}$ (under the assumption that all correspondence relations are one-to-one and that they belong to the subset $\mathcal{F}$).

6.3 Summary

This section has completed the theory of output-drivenness within the OT and HG implementations of constraint-based phonology. Output-drivenness follows from conditions on the faithfulness constraints, namely the FODC$^{\mathrm{OT}}_{\mathrm{comp}}$ and the FODC$^{\mathrm{HG}}_{\mathrm{comp}}$. These conditions are equivalent for faithfulness constraints which satisfy McCarthy's (strengthened) categoricity conjecture, because both conditions can be interpreted as requiring faithfulness constraints to measure phonological distances in compliance with the triangle inequality, formalized as the FTI$_{\mathrm{comp}}$.

7 Conclusions

Idempotency requires any phonotactically licit form to be faithfully realized. Output-drivenness requires any discrepancy between underlying and surface (or output) forms to be driven by the phonotactics. These notions are related to phonological opacity, because their failure corresponds to chain shifts and saltations. Furthermore, the typological structure provided by idempotency and output-drivenness has been shown to have substantial implications for the learnability of phonology. This paper has reviewed (a slight reformulation of) Tesar's (2013) theory of OT output-drivenness and Magri's (to appear) theory of OT idempotency. These analyses provide tight guarantees for OT idempotency and output-drivenness through conditions on the faithfulness constraints, referred to as the FIC$^{\mathrm{OT}}_{\mathrm{comp}}$ and the FODC$^{\mathrm{OT}}_{\mathrm{comp}}$ respectively. This paper has then extended those analyses from OT to HG, obtaining the faithfulness constraint conditions for HG idempotency and output-drivenness referred to as the FIC$^{\mathrm{HG}}_{\mathrm{comp}}$ and the FODC$^{\mathrm{HG}}_{\mathrm{comp}}$.

The four conditions FIC$^{\mathrm{OT}}_{\mathrm{comp}}$, FODC$^{\mathrm{OT}}_{\mathrm{comp}}$, FIC$^{\mathrm{HG}}_{\mathrm{comp}}$, and FODC$^{\mathrm{HG}}_{\mathrm{comp}}$ thus obtained are technical conditions which lack *prima facie* any intuitive interpretation. To provide such an interpretation, the paper has dug deeper into the formal underpinning of the theory of faithfulness. Intuitively, faithfulness constraints measure the phonological distance between underlying and surface forms. It thus makes sense to investigate whether faithfulness constraints satisfy axioms of the abstract notion of distance. The paper has focused on one such axiom, the *triangle inequality*. The main result obtained is that:

(86) $\boxed{\text{The four abstract conditions } \mathrm{FIC}^{\mathrm{OT}}, \mathrm{FODC}^{\mathrm{OT}}, \mathrm{FIC}^{\mathrm{HG}} \text{ and } \mathrm{FODC}^{\mathrm{HG}} \text{ for idempotency and output-drivenness in OT and HG are all equivalent to the faithfulness triangle inequality.}}$

The four conditions can thus all be interpreted as simply requiring the faithfulness constraints to measure phonological distance in compliance with a core axiom of the abstract notion of distance. This equivalence (86) holds for categorical faithfulness constraints in the case of the two OT conditions $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ and $\mathrm{FODC}^{\mathrm{OT}}_{\mathrm{comp}}$, while it holds without restrictions on the nature of the faithfulness constraints in the case of the two HG conditions $\mathrm{FODC}^{\mathrm{HG}}$ and $\mathrm{FIC}^{\mathrm{HG}}$.

This metric interpretation (86) of the four conditions $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$, $\mathrm{FODC}^{\mathrm{OT}}_{\mathrm{comp}}$, $\mathrm{FIC}^{\mathrm{HG}}_{\mathrm{comp}}$, and $\mathrm{FODC}^{\mathrm{HG}}_{\mathrm{comp}}$ for idempotency and output-drivenness has various implications for phonological theory. First, (86) entails that idempotency and output-drivenness do not require stronger constraint conditions in HG than in OT, at least when we restrict ourselves to categorical constraints, as independently conjectured by McCarthy (2003b). Second, (86) entails that the conditions for idempotency and output-drivenness (in either OT or HG) are equivalent when we restrict ourselves to categorical constraints, under the additional assumption that the similarity order which underlies output–drivenness takes into account all of the faithfulness constraints. In other words, the only way to obtain idempotent grammars which fail at output-drivenness is to define output-drivenness relative to a similarity order which is blind to some of the faithfulness constraints. Third, (86) allows the results on which constraints satisfy the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ obtained in Magri (to appear) to be extended in a snap to the other three conditions $\mathrm{FODC}^{\mathrm{OT}}_{\mathrm{comp}}$, $\mathrm{FIC}^{\mathrm{HG}}_{\mathrm{comp}}$, and $\mathrm{FODC}^{\mathrm{HG}}_{\mathrm{comp}}$.

References

Blaho S, Bye P, Krämer M (2007) Freedom of Analysis? Walter de Gruyter

Bolognesi R (1998) The phonology of Campidanian Sardinian: A unitary account of self-organizing structure. PhD thesis, University of Amsterdam, Amsterdam

Buccola B (2013) On the expressivity of optimality theory versus ordered rewrite rules. In: Morrill G, Nederhof M (eds) Proceedings of Formal Grammar 2012 and 2013, Springer, Lecture Notes in Computer Science

Carpenter A (2002) Noncontiguous metathesis and Adjacency. In: Carpenter A, Coetzee A, de Lacy P (eds) Papers in Optimality Theory, vol 2, GLSA, Amherst, MA, pp 1–26

Casali RF (1997) Vowel elision in hiatus contexts: Which vowel goes? Language 73:493–533

Casali RF (1998) Resolving Hiatus. Outstanding dissertations in Linguistics, Garland, New York

Downing LJ (1998) On the prosodic misalignment of onsetless syllables. Natural Language and Linguistic Theory 16:1–52

Downing LJ (2000) Morphological and prosodic constraints on Kinande verbal reduplication. Phonology 17:1–38

Flack K (2007) Inducing functionally grounded constraints. In: Becker M (ed) Papers in Theoretical and Computational Phonology, UMOP (University of Massachusetts Occasional Papers) 36, GLSA, Amherst, MA, pp 13–44

Hayes B (2004) Phonological acquisition in Optimality Theory: The early stages. In: Kager R, Pater J, Zonneveld W (eds) Constraints in Phonological Acquisition, Cambridge University Press, Cambridge, pp 158–203

Heinz J (2005) Reconsidering linearity: Evidence from CV metathesis. In: Alderete J, Han Ch, Kochetov A (eds) Proceedings of WCCFL 24, Cascadilla Press, Somerville, MA, USA, pp 200–208

Keller F (2000) Gradience in grammar. Experimental and computational aspects of degrees of grammaticality. PhD thesis, University of Edinburgh, England

Kubozono H, Ito J, Mester A (2008) Consonant gemination in Japanese loanword phonology. In: 18th International Congress of Linguistics, Seoul

Legendre G, Miyata Y, Smolensky P (1990a) Harmonic Grammar: A formal multi-level connectionist theory of linguistic well-formedness: An application. In: Gernsbacher MA, Derry SJ (eds) Annual conference of the Cognitive Science Society 12, Lawrence Erlbaum Associates, Mahwah, New Jersey, pp 884–891

Legendre G, Miyata Y, Smolensky P (1990b) Harmonic Grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. In: Gernsbacher MA, Derry SJ (eds) Annual conference of the Cognitive Science Society 12, Lawrence Erlbaum, Mahwah, NJ, pp 388–395

Legendre G, Sorace A, Smolensky P (2006) The optimality theory/harmonic grammar connection. In: Smolensky P, Legendre G (eds) The Harminic Mind, MIT Press, Cambridge, MA, pp 903–966

Lombardi L (2001) Why place and voice are different: Constraint interactions and feature faithfulness in optimality theory. In: Lombardi L (ed) Segmental phonology in Optimality Theory: Constraints and Representations, Cambridge University Press, pp 13–45

Łubowicz A (2002) Derived environment effects in Oprimality Theory. Lingua 112:243–280

Łubowicz A (2011) Chain shifts. In: van Oostendorp M, Ewen B Colin Hume, Rice K (eds) Companion to Phonology, Wiley-Blackwell, pp 1717–1735

Magri G (2016a) On the definition of similarity orders in tesar's (2013) theory of output-drivenness, manuscript

Magri G (2016b) Output-drivenness and partial phonological features, manuscript

Magri G (to appear) Idempotency in Optimality Theory. Journal of Linguistics

McCarthy JJ (2002) Comparative markedness (long version), ms. University of Amherst

McCarthy JJ (2003a) Comparative markedness. Theoretical linguistics 29:1–51

McCarthy JJ (2003b) OT constraints are categorical. Phonology 20(1):75–138

McCarthy JJ, Prince A (1995) Faithfulness and reduplicative identity. In: Beckman J, Urbanczyk S, Walsh Dickey L (eds) University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory, GLSA, Amherst, pp 249–384

Moreton E (2004) Non-computable functions in Optimality Theory. In: McCarthy JJ (ed) Optimality Theory in Phonology: A reader, Wiley-Blackwell, Malden: MA, pp 141–163

Moreton E, Smolensky P (2002) Typological consequences of local constraint conjunction. In: Mikkelsen L, Potts C (eds) WCCFL 21: Proceedings of the 21st annual conference of the West Coast Conference on Formal Linguistics, Cascadilla Press, Cambridge, MA, pp 306–319

Prince A (2007) The pursuit of theory. In: de Lacy P (ed) The Cambridge handbook of phonology, Cambridge University Press, pp 33–60

Prince A, Smolensky P (2004) Optimality Theory: Constraint Interaction in generative grammar. Blackwell, Oxford, as Technical Report CU-CS-696-93, Department of Computer Science, University of Colorado at Boulder, and Technical Report TR-2, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, NJ, April 1993. Also available as ROA 537 version

Prince A, Tesar B (2004) Learning phonotactic distributions. In: Kager R, Pater J, Zonneveld W (eds) Constraints in Phonological Acquisition, Cambridge University Press, pp 245–291

Rudin W (1953) Principles of Mathematical Analysis. McGraw-Hill Book Company

Smolensky P (1995) On the internal structure of the constraint component of UG, URL http://roa.rutgers.edu/article/view/87, colloquium presented at the Univ. of California, Los Angeles, April 7, 1995. Handout available as ROA-86 from the Rutgers Optimality Archive

Smolensky P, Legendre G (2006) The Harmonic Mind. MIT Press, Cambridge, MA
Tesar B (2013) Output-Driven Phonology: Theory and Learning. Cambridge Studies in Linguistics
Walker R (1999) Esimbi vowel height shift: implications for faith and markedness, URL http://roa.rutgers.edu/article/view/346, ms, University of Southern California. Available as ROA-336 from the Rutgers Optimality Archive
Wheeler MW (2005) Cluster reduction: Deletion or coalescence? Catalan Journal of Linguistics 4:57–82
White J (2013) Bias in phonological learning: Evidence from saltation. PhD thesis, UCLA

A Proofs

Throughout this appendix, I consider four strings $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$, whose generic segments are denoted by a, b, c, d. For readability, I use statements such as "for every/some segment a" as a shorthand for "for every/some segment a *of the string* $\mathbf{a}$", thus leaving the domain of the quantifier implicit.

A.1 Proof of Proposition 3

Proposition 3 *Assume the candidate set (2) satisfies the transitivity axiom (6) and only contains one-to-one correspondence relations. Consider a faithfulness constraint F which is C-categorical; or I-categorical and O-monotone; or O-categorical and I-monotone. F satisfies the FIC^{OT}_{comp} if and only if it satisfies the FTI_{comp}.* □

Proof As shown in Subsection 3.3.1, the $\mathrm{FTI}_{\mathrm{comp}}$ entails the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ in the general case. To prove the reverse entailment, consider a faithfulness constraint F which satisfies the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ repeated in (87) for any two candidates $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ and $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ and their composition candidate $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$, and let me show that F then satisfies the $\mathrm{FTI}_{\mathrm{comp}}$ repeated in (88).

(87) **If:** $F\big(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}\big) = 0$
Then: $F\big(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}\big) \leq F\big(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}\big)$

(88) $F\big(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}\big) \leq F\big(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}\big) + F\big(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}\big)$

For concreteness, the rest of the proof considers the case where F is I-categorical of order ℓ and O-monotone, so that it satisfies the I-additivity condition repeated in (89); the cases where F is instead C-categorical or O-categorical and I-monotone are treated analogously.

(89) $$F\big(\mathbf{a},\, \mathbf{b},\, \rho_{\mathbf{a},\mathbf{b}}\big) = \sum_{\mathsf{a}_1\cdots\mathsf{a}_\ell \subseteq \mathbf{a}} F\big(\mathsf{a}_1 \cdots \mathsf{a}_\ell,\, \mathbf{b},\, \rho_{\mathbf{a},\mathbf{b}}\restriction_{(\mathsf{a}_1\cdots\mathsf{a}_\ell,\mathbf{b})}\big)$$

The I-additivity condition (89) entails that F assigns zero violations to candidates whose underlying string is shorter than ℓ, as the sum on the right-hand side is empty in this case (there are no subsequences of length ℓ). The $\mathrm{FTI}_{\mathrm{comp}}$

(88) thus trivially holds when its string $\mathbf{a}$ is shorter than ℓ, because its left-hand side $F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$ is equal to zero. From now on, I assume therefore that the string $\mathbf{a}$ has length at least ℓ.

Consider a subsequence $\mathsf{a}_1 \cdots \mathsf{a}_\ell$ of $\mathbf{a}$ of length ℓ. Let $\mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell}$ be the surface correspondent subsequence in $\mathbf{b}$ of the underlying subsequence $\mathsf{a}_1 \cdots \mathsf{a}_\ell$ relative to the correspondence relation $\rho_{\mathbf{a},\mathbf{b}}$ (namely, $\mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell}$ is the subsequence of $\mathbf{b}$ consisting of all and only the segments which are in correspondence with one of the segments $\mathsf{a}_1, \ldots, \mathsf{a}_\ell$). The operations of composition and restriction over correspondence relations commute in the sense of the identity (90): the restriction of the composition correspondence relation $\rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}$ to the pair of strings $(\mathsf{a}_1 \cdots \mathsf{a}_\ell, \mathbf{c})$ coincides with the composition of the restrictions of the relations $\rho_{\mathbf{a},\mathbf{b}}$ and $\rho_{\mathbf{b},\mathbf{c}}$ to the pairs of strings $(\mathsf{a}_1 \cdots \mathsf{a}_\ell, \mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell})$ and $(\mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell}, \mathbf{c})$.

(90) $$\big(\rho_{\mathbf{a},\mathbf{b}} \circ \rho_{\mathbf{b},\mathbf{c}}\big) \restriction_{(\mathsf{a}_1\cdots\mathsf{a}_\ell,\, \mathbf{c})} = \big(\rho_{\mathbf{a},\mathbf{b}} \restriction_{(\mathsf{a}_1\cdots\mathsf{a}_\ell,\, \mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell})}\big) \circ \big(\rho_{\mathbf{b},\mathbf{c}} \restriction_{(\mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell},\, \mathbf{c})}\big)$$

The identity (90) says that the candidate (91c) is the composition of the two candidates (91a) and (91b).

(91) a. $\big(\mathsf{a}_1 \cdots \mathsf{a}_\ell,\ \mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell},\ \rho_{\mathbf{a},\mathbf{b}} \restriction_{(\mathsf{a}_1\cdots\mathsf{a}_\ell,\, \mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell})}\big)$

b. $\big(\mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell},\ \mathbf{c},\ \rho_{\mathbf{b},\mathbf{c}} \restriction_{(\mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell},\, \mathbf{c})}\big)$

c. $\big(\mathsf{a}_1 \cdots \mathsf{a}_\ell,\ \mathbf{c},\ \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}} \restriction_{(\mathsf{a}_1\cdots\mathsf{a}_\ell,\, \mathbf{c})}\big)$

The hypothesis that F satisfies the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ (87) for these two candidates (91a) and (91b) and their composition candidate (91c) becomes:

(92) **If:** $F\big(\mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell},\ \mathbf{c},\ \rho_{\mathbf{b},\mathbf{c}} \restriction_{(\mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell},\, \mathbf{c})}\big) = 0$

Then: $$F\big(\mathsf{a}_1 \cdots \mathsf{a}_\ell,\ \mathbf{c},\ \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}} \restriction_{(\mathsf{a}_1\cdots\mathsf{a}_\ell,\, \mathbf{c})}\big) \leq$$ $$\leq F\big(\mathsf{a}_1 \cdots \mathsf{a}_\ell,\ \mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell},\ \rho_{\mathbf{a},\mathbf{b}} \restriction_{(\mathsf{a}_1\cdots\mathsf{a}_\ell,\, \mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell})}\big)$$

Since F is I-categorical of order ℓ and since the underlying string $\mathsf{a}_1 \cdots \mathsf{a}_\ell$ has length ℓ, the left-hand side of the inequality in the consequent of (92) is equal to either 0 or 1. By reasoning as in Subsection 3.3.2, the $\mathrm{FIC}^{\mathrm{OT}}_{\mathrm{comp}}$ (92) thus entails the $\mathrm{FTI}_{\mathrm{comp}}$ (93).

(93) $$\underbrace{F\big(\mathsf{a}_1 \cdots \mathsf{a}_\ell,\ \mathbf{c},\ \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}} \restriction_{(\mathsf{a}_1\cdots\mathsf{a}_\ell,\, \mathbf{c})}\big)}_{(a)} \leq$$ $$\leq \underbrace{F\big(\mathsf{a}_1 \cdots \mathsf{a}_\ell,\ \mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell},\ \rho_{\mathbf{a},\mathbf{b}} \restriction_{(\mathsf{a}_1\cdots\mathsf{a}_\ell,\, \mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell})}\big)}_{(b)} + \underbrace{F\big(\mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell},\ \mathbf{c},\ \rho_{\mathbf{b},\mathbf{c}} \restriction_{(\mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell},\mathbf{c})}\big)}_{(c)}$$

The rest of the proof obtains the $\mathrm{FTI}_{\mathrm{comp}}$ (88) by summing the inequality (93) over all subsequences $\mathsf{a}_1 \cdots \mathsf{a}_\ell$ of length ℓ of the underlying string $\mathbf{a}$.

To start, the definition of I-additivity of order ℓ applied to the composition candidate $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$ immediately yields the expression (94) for the sum of the terms (93a) over all subsequences $\mathsf{a}_1 \cdots \mathsf{a}_\ell$ of the underlying string $\mathbf{a}$.

(94) $$\sum_{\mathsf{a}_1\cdots\mathsf{a}_\ell \subseteq \mathbf{a}} \underbrace{F\big(\mathsf{a}_1 \cdots \mathsf{a}_\ell,\ \mathbf{c},\ \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}} \restriction_{(\mathsf{a}_1\cdots\mathsf{a}_\ell,\, \mathbf{c})}\big)}_{(93a)} = F\big(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}\big)$$

The sum of the terms (93b) over all subsequences $\mathsf{a}_1 \cdots \mathsf{a}_\ell$ of the underlying string $\mathbf{a}$ can be upper bounded as in (95). In step (95a), I have used the hypothesis that F is O-monotone (together with the obvious fact that $\mathbf{b}_{\mathsf{a}_1 \cdots \mathsf{a}_\ell}$ is a subsequence of $\mathbf{b}$). Step (95b) follows from the fact that the restriction of $\rho_{\mathbf{a},\mathbf{b}}$ to the string pair $(\mathsf{a}_1 \cdots \mathsf{a}_\ell,\, \mathbf{b}_{\mathsf{a}_1 \cdots \mathsf{a}_\ell})$ is identical to its restriction to the string pair $(\mathsf{a}_1 \cdots \mathsf{a}_\ell,\, \mathbf{b})$, because $\mathbf{b}_{\mathsf{a}_1 \cdots \mathsf{a}_\ell}$ is the subsequence of $\mathbf{b}$ consisting of those segments which are in correspondence with one of the segments $\mathsf{a}_1, \ldots, \mathsf{a}_\ell$. Step (95c) follows again from the hypothesis that F is I-additive of order ℓ.

(95)
$$\sum_{\mathsf{a}_1 \cdots \mathsf{a}_\ell \subseteq \mathbf{a}} \underbrace{F\big(\mathsf{a}_1 \cdots \mathsf{a}_\ell,\, \mathbf{b}_{\mathsf{a}_1 \cdots \mathsf{a}_\ell},\, \rho_{\mathbf{a},\mathbf{b}}\restriction_{(\mathsf{a}_1 \cdots \mathsf{a}_\ell,\, \mathbf{b}_{\mathsf{a}_1 \cdots \mathsf{a}_\ell})}\big)}_{(93b)} \leq$$
$$\overset{(a)}{\leq} \sum_{\mathsf{a}_1 \cdots \mathsf{a}_\ell \subseteq \mathbf{a}} F\big(\mathsf{a}_1 \cdots \mathsf{a}_\ell,\, \mathbf{b},\, \rho_{\mathbf{a},\mathbf{b}}\restriction_{(\mathsf{a}_1 \cdots \mathsf{a}_\ell,\, \mathbf{b}_{\mathsf{a}_1 \cdots \mathsf{a}_\ell})}\big)$$
$$\overset{(b)}{=} \sum_{\mathsf{a}_1 \cdots \mathsf{a}_\ell \subseteq \mathbf{a}} F\big(\mathsf{a}_1 \cdots \mathsf{a}_\ell,\, \mathbf{b},\, \rho_{\mathbf{a},\mathbf{b}}\restriction_{(\mathsf{a}_1 \cdots \mathsf{a}_\ell,\, \mathbf{b})}\big)$$
$$\overset{(c)}{=} F\big(\mathbf{a},\, \mathbf{b},\, \rho_{\mathbf{a},\mathbf{b}}\big)$$

Finally, let me bound the sum of the terms (93c) over all subsequences $\mathsf{a}_1 \cdots \mathsf{a}_\ell$ of the underlying string $\mathbf{a}$. To this end, I note that the implication (96) holds for any two subsequences $\mathsf{a}_1 \cdots \mathsf{a}_\ell$ and $\widehat{\mathsf{a}}_1 \cdots \widehat{\mathsf{a}}_\ell$ and their surface correspondent subsequences $\mathbf{b}_{\mathsf{a}_1 \cdots \mathsf{a}_\ell}$ and $\mathbf{b}_{\widehat{\mathsf{a}}_1 \cdots \widehat{\mathsf{a}}_\ell}$ of $\mathbf{b}$.

(96) **If:** $\mathsf{a}_1 \cdots \mathsf{a}_\ell \neq \widehat{\mathsf{a}}_1 \cdots \widehat{\mathsf{a}}_\ell$ and $\mathbf{b}_{\mathsf{a}_1 \cdots \mathsf{a}_\ell}$ has length at least ℓ
Then: $\mathbf{b}_{\mathsf{a}_1 \cdots \mathsf{a}_\ell} \neq \mathbf{b}_{\widehat{\mathsf{a}}_1 \cdots \widehat{\mathsf{a}}_\ell}$

In fact, assume by contradiction that the antecedent holds but the consequent fails. Since the surface correspondent string $\mathbf{b}_{\mathsf{a}_1 \cdots \mathsf{a}_\ell}$ has length at least ℓ and since the correspondence relation $\rho_{\mathbf{a},\mathbf{b}}$ cannot break any underlying segment into two or more surface segments (because it is one-to-one), each underlying segment a_i of $\mathsf{a}_1 \cdots \mathsf{a}_\ell$ must have a surface correspondent in $\mathbf{b}_{\mathsf{a}_1 \cdots \mathsf{a}_\ell}$. The hypothesis $\mathsf{a}_1 \cdots \mathsf{a}_\ell \neq \widehat{\mathsf{a}}_1 \cdots \widehat{\mathsf{a}}_\ell$ means that there exists at least one segment a_i which belongs to $\mathsf{a}_1 \cdots \mathsf{a}_\ell$ but not to $\widehat{\mathsf{a}}_1 \cdots \widehat{\mathsf{a}}_\ell$. Let b be the surface correspondent of a_i in $\mathbf{b}_{\mathsf{a}_1 \cdots \mathsf{a}_\ell}$. Because of the contradictory assumption that the consequent of (96) fails, $\mathbf{b}_{\mathsf{a}_1 \cdots \mathsf{a}_\ell} = \mathbf{b}_{\widehat{\mathsf{a}}_1 \cdots \widehat{\mathsf{a}}_\ell}$. This means that b also belongs to $\mathbf{b}_{\widehat{\mathsf{a}}_1 \cdots \widehat{\mathsf{a}}_\ell}$, namely must correspond to some segment $\widehat{\mathsf{a}}_j$ of $\widehat{\mathsf{a}}_1 \cdots \widehat{\mathsf{a}}_\ell$. Since a_i does not belong to $\widehat{\mathsf{a}}_1 \cdots \widehat{\mathsf{a}}_\ell$, then a_i and $\widehat{\mathsf{a}}_j$ must be different. The conclusion that both $(\mathsf{a}_i, \mathsf{b})$ and $(\widehat{\mathsf{a}}_j, \mathsf{b})$ belong to $\rho_{\mathbf{a},\mathbf{b}}$ despite the fact that $\mathsf{a}_i \neq \widehat{\mathsf{a}}_j$ contradicts the hypothesis that $\rho_{\mathbf{a},\mathbf{b}}$ does not coalesce any two underlying segments.

Let me now go back to the goal of bounding the sum of the terms (93c) over all subsequences $\mathsf{a}_1 \cdots \mathsf{a}_\ell$ of the underlying string $\mathbf{a}$. Since $\mathsf{a}_1 \cdots \mathsf{a}_\ell$ has length ℓ and since the correspondence relation $\rho_{\mathbf{a},\mathbf{b}}$ cannot break any underlying segment (because it is one-to-one), the surface correspondent string $\mathbf{b}_{\mathsf{a}_1 \cdots \mathsf{a}_\ell}$ has length ℓ or smaller. If $\mathbf{b}_{\mathsf{a}_1 \cdots \mathsf{a}_\ell}$ has length smaller than ℓ, the corresponding term (93c) is null, because F is I-additive of order ℓ and thus assigns zero violations to candidates whose underlying string is shorter than ℓ, as noted at the beginning. The sum can thus be restricted to candidates whose underlying form $\mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell}$ has length exactly ℓ, as in step (97a). Condition (96) says that the mapping from the subsequences $\mathsf{a}_1\cdots\mathsf{a}_\ell$ to the corresponding surface subsequences $\mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell}$ (of length ℓ) is an injection, thus guaranteeing step (97b). Step (97c) follows again from the hypothesis that F is I-additive of order ℓ.

(97)
$$\begin{aligned}
&\sum_{\mathsf{a}_1\cdots\mathsf{a}_\ell\subseteq\mathbf{a}} \underbrace{F\big(\mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell},\ \mathbf{c},\ \rho_{\mathbf{b},\mathbf{c}}\restriction_{(\mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell},\,\mathbf{c})}\big)}_{(93c)}\\
&\overset{(a)}{=} \sum_{\substack{\mathsf{a}_1\cdots\mathsf{a}_\ell\subseteq\mathbf{a}\\ \text{s.t. } \mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell} \text{ has length } \ell}} F\big(\mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell},\ \mathbf{c},\ \rho_{\mathbf{b},\mathbf{c}}\restriction_{(\mathbf{b}_{\mathsf{a}_1\cdots\mathsf{a}_\ell},\,\mathbf{c})}\big)\\
&\overset{(b)}{\leq} \sum_{\mathsf{b}_1\cdots\mathsf{b}_\ell\subseteq\mathbf{b}} F\big(\mathsf{b}_1\cdots\mathsf{b}_\ell,\ \mathbf{c},\ \rho_{\mathbf{b},\mathbf{c}}\restriction_{(\mathsf{b}_1\cdots\mathsf{b}_\ell,\,\mathbf{c})}\big)\\
&\overset{(c)}{=} F\big(\mathbf{b},\ \mathbf{c},\ \rho_{\mathbf{b},\mathbf{c}}\big)
\end{aligned}$$

The $\mathrm{FTI}_{\mathrm{comp}}$ (88) follows by summing the inequality (93) over all subsequences $\mathsf{a}_1\cdots\mathsf{a}_\ell$ of length ℓ of the string $\mathbf{a}$, using the three expressions (94), (95), and (97) for the sums over the three terms (93a), (93b), and (93c). □

A.2 Proof of Proposition 5

Proposition 5 *Assume that, for any two candidates $(\boldsymbol{a}, \boldsymbol{b}, \rho_{\boldsymbol{a},\boldsymbol{b}})$ and $(\boldsymbol{b}, \boldsymbol{c}, \rho_{\boldsymbol{b},\boldsymbol{c}})$, the candidate set also contains a candidate $(\boldsymbol{a}, \boldsymbol{c}, \rho_{\boldsymbol{a},\boldsymbol{c}})$ such that the FIC^{HG} repeated in (98) holds for any faithfulness constraint F in the constraint set.*

(98) *For every choice of the constant $\xi \geq 0$:*

If: $F\big(\boldsymbol{b}, \mathbf{c}, \rho_{\boldsymbol{b},\boldsymbol{c}}\big) \leq \xi$

Then: $F\big(\boldsymbol{a}, \boldsymbol{c}, \rho_{\boldsymbol{a},\boldsymbol{c}}\big) \leq F\big(\boldsymbol{a}, \boldsymbol{b}, \rho_{\boldsymbol{a},\boldsymbol{b}}\big) + \xi$

Then, the HG grammar corresponding to any weighting of the constraint set is idempotent, no matter what the markedness constraints look like. □

Proof Suppose that the HG grammar $G_{\boldsymbol{\theta}}$ corresponding to some weighting $\boldsymbol{\theta}$ fails at the idempotency implication (32) for some candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$, as stated in (99): $G_{\boldsymbol{\theta}}$ maps the underlying form $\mathbf{a}$ to $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$, as required by the antecedent of the idempotency implication; but it fails to map the underlying form $\mathbf{b}$ to the identity candidate $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$, as required by the consequent.

(99) $G_{\boldsymbol{\theta}}$ fails at idempotency on a candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ if and only if:

a. $G_{\boldsymbol{\theta}}(\mathbf{a}) = (\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$;

b. $G_{\boldsymbol{\theta}}(\mathbf{b}) \neq (\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$.

Condition (99b) means that the grammar $G_{\boldsymbol{\theta}}$ maps the underlying form $\mathbf{b}$ to a candidate $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ different from $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$. This means that either the two strings $\mathbf{b}$ and $\mathbf{c}$ differ or else $\mathbf{b}$ and $\mathbf{c}$ coincide but the two correspondence relations $\rho_{\mathbf{b},\mathbf{c}}$ and $\mathbb{I}_{\mathbf{b},\mathbf{b}}$ differ. The latter option is impossible, because the candidate $(\mathbf{b}, \mathbf{b}, \rho_{\mathbf{b},\mathbf{b}})$ with $\rho_{\mathbf{b},\mathbf{b}} \neq \mathbb{I}_{\mathbf{b},\mathbf{b}}$ is harmonically bounded by $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$: faithfulness constraints cannot prefer the former candidate, by (8); and markedness constraints cannot distinguish between the two candidates, by (7). The two strings $\mathbf{b}$ and $\mathbf{c}$ must therefore differ and condition (99) becomes (100).

(100) $G_{\boldsymbol{\theta}}$ fails at idempotency on a candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ if and only if there exists a candidate $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ with $\mathbf{b} \neq \mathbf{c}$ such that:

a. $G_{\boldsymbol{\theta}}(\mathbf{a}) = (\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$;

b. $G_{\boldsymbol{\theta}}(\mathbf{b}) = (\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$.

By assumption, the two candidates $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ and $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ come with a companion candidate $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$. The "if-and-only-if" statement (100) can thus be weakened to the "if" statement (101). In fact, if the grammar $G_{\boldsymbol{\theta}}$ maps the underlying form $\mathbf{a}$ to the candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ as stated in (100a), the weights $\boldsymbol{\theta}$ prefer this candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ to the candidate $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$, as stated in (101a). Furthermore, if the grammar $G_{\boldsymbol{\theta}}$ maps the underlying form $\mathbf{b}$ to the candidate $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ as stated in (101b), the weights $\boldsymbol{\theta}$ prefer this candidate $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ to the identity candidate $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$, as stated in (101b).

(101) If $G_{\boldsymbol{\theta}}$ fails at idempotency on a candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$, there exists some candidate $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ with $\mathbf{b} \neq \mathbf{c}$ such that:

a. $\boldsymbol{\theta}$ prefers $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ to $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$,

b. $\boldsymbol{\theta}$ prefers $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ to $(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$.

Condition (101) can be made explicit as in (102) in terms of the number of constraint violations. These sums run over a generic markedness constraint M with weight θ_M and a generic faithfulness constraint F with weight θ_F. The faithfulness constraints do not appear on the right-hand side of (102b) because $F(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}}) = 0$ for every faithfulness constraint F, by (8).

(102) If $G_{\boldsymbol{\theta}}$ fails at idempotency on a candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$, there exists some candidate $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ with $\mathbf{b} \neq \mathbf{c}$ such that:

a. $$\sum_M \theta_M M(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) + \sum_F \theta_F F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) < \\ < \sum_M \theta_M M(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) + \sum_F \theta_F F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$$

b. $$\sum_M \theta_M M(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) + \sum_F \theta_F F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) < \sum_M \theta_M M(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}})$$

Let the constant ξ be defined as $\xi = \sum_M \theta_M M(\mathbf{b}, \mathbf{b}, \mathbb{I}_{\mathbf{b},\mathbf{b}}) - \sum_M \theta_M M(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$. Since markedness constraints are blind to underlying forms by (7), then also $\xi = \sum_M \theta_M M(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) - \sum_M \theta_M M(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$. Condition (102) thus becomes:

(103) If $G_{\boldsymbol{\theta}}$ fails at idempotency on a candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$, there exists some candidate $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ with $\mathbf{b} \neq \mathbf{c}$ such that:

a. $\sum_F \theta_F F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) > \sum_F \theta_F F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) + \xi$

b. $\sum_F \theta_F F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) < \xi$

In conclusion, idempotency holds for the HG grammar corresponding to any weighting of the constraint set provided the two conditions (103a) and (103b) can never be satisfied both, no matter the choice of the weights θ_F and the constant ξ. In other words, it suffices to assume that, for every two candidates $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ and $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$, there exists some candidate $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$ such that:

(104) For every choice of the constant $\xi \geq 0$, for every choice of the weights θ_F:

If: $\sum_F \theta_F F\big(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}\big) < \xi$

Then: $\sum_F \theta_F F\big(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}\big) \leq \sum_F \theta_F F\big(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}\big) + \xi$

To conclude the proof, I need to show that (104) is equivalent to the $\mathrm{FIC}^{\mathrm{HG}}$ (98). To start, let me show that (98) entails (104). In fact, suppose that the antecedent of the implication (104) holds. For every faithfulness constraint F, let ξ_F be defined as in (105a). The antecedent of the implication (104) can thus be restated as in (105b). I can assume without loss of generality that the weights θ_F are all different from zero. The position (105a) thus entails (105c). Since the implication (98) holds by hypothesis, (105c) entails (105d). The consequent of the implication (104) thus follows from (105b) by taking the weighted average of the inequalities (105d) over all faithfulness constraints.

(105) a. $\xi_F = \theta_F F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$

b. $\sum_F \xi_F < \xi$

c. $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq \xi_F / \theta_F$

d. $F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}) \leq F(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}}) + \xi_F / \theta_F$

Let me now show that (104) vice versa entails (98). In fact, suppose that the antecedent of the implication (98) holds, namely that $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq \xi$. Let me distinguish two cases, depending on whether ξ is an integer or not. To start, assume that ξ is not an integer. The assumption $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq \xi$ (with the loose inequality) is thus equivalent to the assumption $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) < \xi$ (with the strict inequality), because constraint violations are integers. The antecedent of the implication (104) thus holds with all the weights set equal to zero but for the weight θ_F corresponding to the faithfulness constraint F considered, which is equal to 1. The consequent of the implication (104) must therefore hold as well, which is in turn identical to the consequent of the implication (98) with this special choice of the weights. If instead the antecedent of (98) holds with ξ equal to an integer, let $\widehat{\xi} = \xi + 1/2$. By reasoning as above, I conclude that the consequent of the implication (98) holds for $\widehat{\xi}$. Since constraint violations are integers, the latter entails in turn that the consequent of the implication (98) holds for ξ. □

A.3 Proof of Proposition 12

Proposition 12 *Assume that, for any two candidates $(\boldsymbol{a}, \boldsymbol{d}, \rho_{\boldsymbol{a},\boldsymbol{d}})$ and $(\boldsymbol{b}, \boldsymbol{d}, \rho_{\boldsymbol{b},\boldsymbol{d}})$ such that $(\boldsymbol{a}, \boldsymbol{d}, \rho_{\boldsymbol{a},\boldsymbol{d}}) \leq_{sim} (\boldsymbol{b}, \boldsymbol{d}, \rho_{\boldsymbol{b},\boldsymbol{d}})$, for every candidate $(\boldsymbol{b}, \boldsymbol{c}, \rho_{\boldsymbol{b},\boldsymbol{c}})$ different from $(\boldsymbol{b}, \boldsymbol{d}, \rho_{\boldsymbol{b},\boldsymbol{d}})$, the candidate set also contains a candidate $(\boldsymbol{a}, \boldsymbol{c}, \rho_{\boldsymbol{a},\boldsymbol{c}})$ different from $(\boldsymbol{a}, \boldsymbol{d}, \rho_{\boldsymbol{a},\boldsymbol{d}})$ such that the FODC^{HG} repeated in (106) holds for any faithfulness constraint F.*

(106) For every choice of the constant ξ (with no restrictions on its sign):

If: $F\big(\boldsymbol{b}, \boldsymbol{c}, \rho_{\boldsymbol{b},\boldsymbol{c}}\big) \leq F\big(\boldsymbol{b}, \boldsymbol{d}, \rho_{\boldsymbol{b},\boldsymbol{d}}\big) + \xi$

Then: $F\big(\boldsymbol{a}, \boldsymbol{c}, \rho_{\boldsymbol{a},\boldsymbol{c}}\big) \leq F\big(\boldsymbol{a}, \boldsymbol{d}, \rho_{\boldsymbol{a},\boldsymbol{d}}\big) + \xi$

Then, the HG grammar corresponding to any weighting of the constraint set is output-driven relative to the similarity order $\leq_{sim}$. □

Proof The proof is similar to the proof of Proposition 5 in Appendix A.2. Suppose that the HG grammar $G_{\boldsymbol{\theta}}$ corresponding to some weighting $\boldsymbol{\theta}$ fails at the output-drivenness implication (51) for two candidates $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ and $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$, as stated in (107): the grammar $G_{\boldsymbol{\theta}}$ maps the underlying form $\mathbf{a}$ to the candidate $(\mathbf{a}, \mathbf{b}, \rho_{\mathbf{a},\mathbf{b}})$ with less internal similarity, as required by the antecedent of (51); but it fails to map the underlying form $\mathbf{b}$ to the candidate $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ with more internal similarity, as required by the consequent of (51).

(107) $G_{\boldsymbol{\theta}}$ fails at output-drivenness on candidates $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ iff there exists a candidate $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ different from $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ such that:

a. $G_{\boldsymbol{\theta}}(\mathbf{a}) = (\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$;

b. $G_{\boldsymbol{\theta}}(\mathbf{b}) = (\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \neq (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$.

By assumption, the candidate $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ different from $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ comes with a companion candidate $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$ different from $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$. The "if-and-only-if" statement (107) can thus be weakened into the "if" statement (108).

(108) If $G_{\boldsymbol{\theta}}$ fails at output-drivenness on candidates $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$, there exists a candidate $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ different from $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ such that:

a. $\boldsymbol{\theta}$ prefers $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}})$ to $(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$,

b. $\boldsymbol{\theta}$ prefers $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ to $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$.

Condition (108) can be made explicit as in (109) in terms of the numbers of constraint violations. The sums run over a generic markedness constraint M with weight θ_M and a generic faithfulness constraint F with weight θ_F.

(109) If $G_{\boldsymbol{\theta}}$ fails at output-drivenness on candidates $(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}} (\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$, there exists a candidate $(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}})$ different from $(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$ such that:

$$\text{a.}\quad \sum_M \theta_M M(\mathbf{a},\mathbf{d},\rho_{\mathbf{a},\mathbf{d}}) + \sum_F \theta_F F(\mathbf{a},\mathbf{d},\rho_{\mathbf{a},\mathbf{d}}) < \\ < \sum_M \theta_M M(\mathbf{a},\mathbf{c},\rho_{\mathbf{a},\mathbf{c}}) + \sum_F \theta_F F(\mathbf{a},\mathbf{c},\rho_{\mathbf{a},\mathbf{c}})$$

$$\text{b.}\quad \sum_M \theta_M M(\mathbf{b},\mathbf{c},\rho_{\mathbf{b},\mathbf{c}}) + \sum_F \theta_F F(\mathbf{b},\mathbf{c},\rho_{\mathbf{b},\mathbf{c}}) < \\ < \sum_M \theta_M M(\mathbf{b},\mathbf{d},\rho_{\mathbf{b},\mathbf{d}}) + \sum_F \theta_F F(\mathbf{b},\mathbf{d},\rho_{\mathbf{b},\mathbf{d}})$$

Taking advantage of the fact that markedness constraints are blind to underlying forms by (7), condition (109) can be rewritten as in (110) with the position $\xi = \sum_M \theta_M M(\mathbf{a},\mathbf{d},\rho_{\mathbf{a},\mathbf{d}}) - \sum_M \theta_M M(\mathbf{a},\mathbf{c},\rho_{\mathbf{a},\mathbf{c}}) = \sum_M \theta_M M(\mathbf{b},\mathbf{d},\rho_{\mathbf{b},\mathbf{d}}) - \sum_M \theta_M M(\mathbf{b},\mathbf{c},\rho_{\mathbf{b},\mathbf{c}})$.

(110) If $G_{\boldsymbol{\theta}}$ fails at output-drivenness on candidates $(\mathbf{a},\mathbf{d},\rho_{\mathbf{a},\mathbf{d}}) \leq_{\text{sim}} (\mathbf{b},\mathbf{d},\rho_{\mathbf{b},\mathbf{d}})$, there exists a candidate $(\mathbf{b},\mathbf{c},\rho_{\mathbf{b},\mathbf{c}})$ different from $(\mathbf{b},\mathbf{d},\rho_{\mathbf{b},\mathbf{d}})$ such that:

$$\text{a.}\quad \sum_F \theta_F F(\mathbf{a},\mathbf{c},\rho_{\mathbf{a},\mathbf{c}}) > \sum_F \theta_F F(\mathbf{a},\mathbf{d},\rho_{\mathbf{a},\mathbf{d}}) + \xi$$

$$\text{b.}\quad \sum_F \theta_F F(\mathbf{b},\mathbf{c},\rho_{\mathbf{b},\mathbf{c}}) < \sum_F \theta_F F(\mathbf{b},\mathbf{d},\rho_{\mathbf{b},\mathbf{d}}) + \xi$$

In conclusion, output-drivenness holds for the HG grammar corresponding to any constraint weighting provided the two conditions (110a) and (110b) can never be satisfied both, no matter the choice of the weights θ_F and the constant ξ. In other words, it suffices to assume that for every candidate $(\mathbf{b},\mathbf{c},\rho_{\mathbf{b},\mathbf{c}})$ different from $(\mathbf{b},\mathbf{d},\rho_{\mathbf{b},\mathbf{d}})$ there exists a candidate $(\mathbf{a},\mathbf{c},\rho_{\mathbf{a},\mathbf{c}})$ different from $(\mathbf{a},\mathbf{d},\rho_{\mathbf{a},\mathbf{d}})$ such that:

(111) For every choice of the constant ξ, for every choice of the weights θ_F:

$$\textbf{If:}\quad \sum_F \theta_F F\big(\mathbf{b},\mathbf{c},\rho_{\mathbf{b},\mathbf{c}}\big) < \sum_F \theta_F F\big(\mathbf{b},\mathbf{d},\rho_{\mathbf{b},\mathbf{d}}\big) + \xi$$

$$\textbf{Then:}\quad \sum_F \theta_F F\big(\mathbf{a},\mathbf{c},\rho_{\mathbf{a},\mathbf{c}}\big) \leq \sum_F \theta_F F\big(\mathbf{a},\mathbf{d},\rho_{\mathbf{a},\mathbf{d}}\big) + \xi$$

To conclude the proof, condition (111) can be shown to be equivalent to the FODC$^{\text{HG}}$ (106) by reasoning as at the end of the proof of Proposition 5 to show that condition (104) is equivalent to the FIC$^{\text{HG}}$ (98). □

A.4 Proof of Proposition 13

Proposition 13 *The two FODC^{OT} implications repeated in (112)*

(112) *a.*

$$\textbf{If:}\quad F\big(\boldsymbol{a},\boldsymbol{d},\rho_{\boldsymbol{a},\boldsymbol{d}}\big) < F\big(\boldsymbol{a},\boldsymbol{c},\rho_{\boldsymbol{a},\boldsymbol{c}}\big)$$
$$\textbf{Then:}\quad F\big(\boldsymbol{b},\boldsymbol{d},\rho_{\boldsymbol{b},\boldsymbol{d}}\big) < F\big(\boldsymbol{b},\boldsymbol{c},\rho_{\boldsymbol{b},\boldsymbol{c}}\big)$$

b.

$$\textbf{If:}\quad F\big(\boldsymbol{b},\boldsymbol{c},\rho_{\boldsymbol{b},\boldsymbol{c}}\big) < F\big(\boldsymbol{b},\boldsymbol{d},\rho_{\boldsymbol{b},\boldsymbol{d}}\big)$$
$$\textbf{Then:}\quad F\big(\boldsymbol{a},\boldsymbol{c},\rho_{\boldsymbol{a},\boldsymbol{c}}\big) < F\big(\boldsymbol{a},\boldsymbol{d},\rho_{\boldsymbol{a},\boldsymbol{d}}\big)$$

are jointly equivalent to the following condition:

(113) *For every choice of the constant ξ between -1 and $+1$ (both excluded):*
If: $F\big(\boldsymbol{b}, \boldsymbol{c}, \rho_{\boldsymbol{b},\boldsymbol{c}}\big) \leq F\big(\boldsymbol{b}, \boldsymbol{d}, \rho_{\boldsymbol{b},\boldsymbol{d}}\big) + \xi$
Then: $F\big(\boldsymbol{a}, \boldsymbol{c}, \rho_{\boldsymbol{a},\boldsymbol{c}}\big) \leq F\big(\boldsymbol{a}, \boldsymbol{d}, \rho_{\boldsymbol{a},\boldsymbol{d}}\big) + \xi$ □

Proof Let me show that the two FODC^{OT} implications (112) jointly entail the implication (113). Thus, assume that the antecedent of the latter implication holds for some ξ. I distinguish two cases, depending on whether ξ is (strictly) smaller than 0 or not. Let me start with the former case, stated in (114a).

(114) a. $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) + \xi$ with $-1 < \xi < 0$
b. $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) < F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$
c. $F\big(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}\big) < F\big(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}\big)$
d. $F\big(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}}\big) \leq F\big(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}\big) + \xi$

Since ξ is strictly negative, (114a) entails the strict inequality (114b). The latter in turn coincides with the antecedent of the second FODC^{OT} implication (112b), which therefore ensures that its consequent holds as well, repeated in (114c). Since ξ is larger than -1 and constraint violations are integers, (114c) in turn entails (114d), which is the desired consequent of the implication (113). Note that this reasoning has used only the second FODC^{OT} implication (112b).

Consider next the complementary case where the antecedent of the implication (113) holds with a nonnegative ξ, as stated in (115a).

(115) a. $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) + \xi$ with $0 \leq \xi < +1$
b. $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \leq F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$
c. $F\big(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}\big) \geq F\big(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}\big)$
d. $F\big(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}\big) + \xi \geq F\big(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}}\big)$

Since ξ is smaller than $+1$ and constraint violations are integers, (114a) entails (114b). The latter in turns says that the consequent of the first FODC^{OT} implication (112a) fails. The antecedent must therefore fail as well, as stated in (114c). Since ξ is nonnegative, the latter in turn entails (114d), which is the desired consequent of the implication (113). Note that this reasoning has used only the first FODC^{OT} implication (112a).

Next, let me show that condition (113) with $0 < \xi < +1$ in turn entails the first FODC^{OT} implication (112a). In fact, suppose that the antecedent of the latter implication holds, namely that $F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) < F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{c}})$. Since ξ is smaller than $+1$ and constraint violations are integers, the latter entails that $F(\mathbf{a}, \mathbf{d}, \rho_{\mathbf{a},\mathbf{d}}) + \xi < F(\mathbf{a}, \mathbf{c}, \rho_{\mathbf{a},\mathbf{b}}\rho_{\mathbf{b},\mathbf{c}})$. The consequent of the implication (113) thus fails. Its antecedent must therefore fail as well, namely $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) \geq F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}}) + \xi$. The latter entails that $F(\mathbf{b}, \mathbf{c}, \rho_{\mathbf{b},\mathbf{c}}) > F(\mathbf{b}, \mathbf{d}, \rho_{\mathbf{b},\mathbf{d}})$, establishing the consequent of the first $\text{FODC}^{\text{OT}}_{\text{comp}}$ implication (112a). An analogous reasoning shows that condition (113) with $-1 < \xi < 0$ entails the second $\text{FODC}^{\text{OT}}_{\text{comp}}$ implication (112b). □